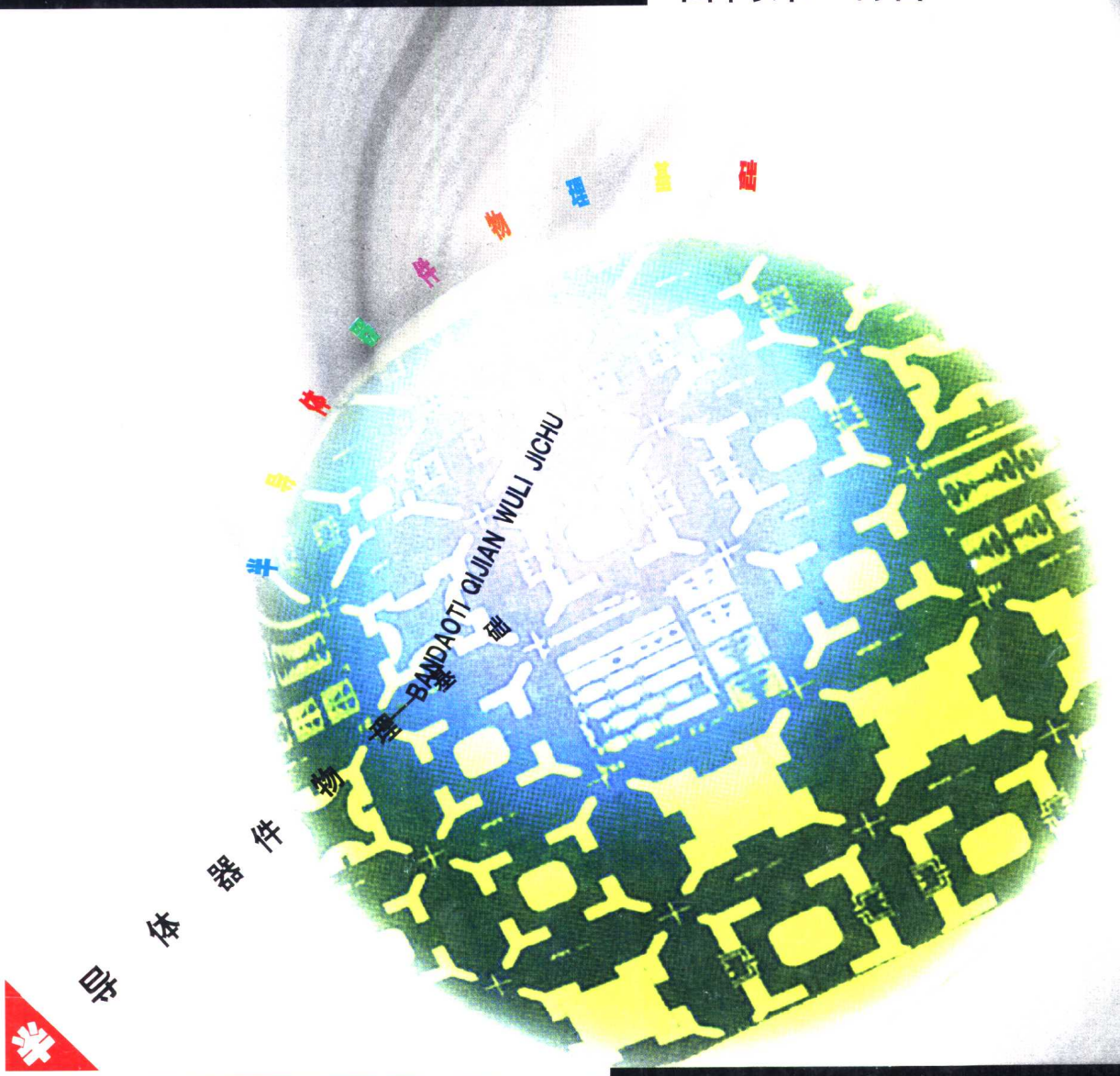
半导体器件物理基础

BANDAOTI QIJIAN WULI JICHU

曾树荣 编著



北京大学出版社

BANDAOTI QIJIAN WULI JICHU

半导体器件物理基础

责任编辑＼沈承凤
封面设计＼张　虹

ISBN 7-301-05456-4
9 787301 054567 >
ISBN 7-301-05456-4/TP・0648
定价：22.00元

半导体器件物理基础

曾树荣　编著

北京大学出版社
北　京

内 容 简 介

本书内容大体可分为两个部分。前两章为第一部分，介绍学习半导体器件必需的知识，包括半导体基本知识和p-n结理论；其余各章为第二部分，阐述半导体器件的基本原理及双极型晶体管、化合物半导体场效应晶体管、MOS器件、有源微波二极管和光电器件等半导体器件。

本书简明扼要，既着重于基本的半导体器件及其基本知识，也涉及了半导体器件领域内的新进展和新概念，可作为电子信息专业特别是微电子学专业的教材或参考书，分章节供本科生和研究生使用；也可供有关研究人员参考。

图书在版编目(CIP)数据

半导体器件物理基础/曾树荣编著. －北京：北京大学出版社，2002.2
ISBN 7-301-05456-4

Ⅰ.半… Ⅱ.曾… Ⅲ.半导体器件-半导体物理 Ⅳ.TN303

中国版本图书馆 CIP 数据核字(2002)第 001726 号

书　　名：半导体器件物理基础
著作责任者：曾树荣
责 任 编 辑：沈承凤
标 准 书 号：ISBN 7-301-05456-4/TP·0648
出 版 者：北京大学出版社
地　　址：北京市海淀区中关村北京大学校内　100871
网　　址：http://cbs.pku.edu.cn
电　　话：出版部 62752015　发行部 62754140　编辑部 62752032
排 版 者：兴盛达打字服务社　62549189
电 子 信 箱：zpup@pup.pku.edu.cn
印 刷 者：北京神剑印刷厂(原国防科工委印刷厂)
发 行 者：北京大学出版社
经 销 者：新华书店
787 毫米×1092 毫米　16 开本　15.875 印张　395 千字
2002 年 2 月第 1 版　2002 年 2 月第 1 次印刷
定　　价：22.00 元

前　言

人类已经进入21世纪，这将是社会高度信息化的世纪。计算机网络和通信技术的进步是信息化发展的标志。今天，无论何时何地，人们都可以高速交换信息，使区域性很强的政治、经济、文化国际化。但是，如果没有以晶体管为基础的微电子学的发展就谈不上个人计算机的高性能和多功能，没有以激光器和光探测器为主体的光电子学的发展就谈不上信息的高速传输，也就是说不可能有现代信息技术。本书介绍微电子学和光电子学领域主要半导体器件的基本原理，供将要从事或者正在从事电子信息专业特别是微电子学专业的人员阅读。

微电子学的主体是在以晶体管、二极管为单元的各种半导体器件的基础上构成的集成电路。1947年半导体三极管(双极晶体管)诞生之后，半个世纪的时间内，半导体器件的理论和制造技术飞速发展。1958年集成电路问世，标志着微电子学时代开始；1968年MOS大规模集成电路产业化，从此集成电路的集成度一直以每3年4倍的惊人速度发展。今天，大部分的半导体器件是用硅做的，硅集成电路占据主导地位。就高速和功率应用而言，双极晶体管一直领先；但MOS场效应晶体管(MOSFET)易于小型化和功耗很小，在集成度方面占有明显的优势，所以MOS器件集成度的提高实际上代表了当今微电子学的进步。但是，随着电子装备向更高级发展，硅器件在许多领域不能适应，化合物半导体电子器件及集成电路成为重要的发展方向。1966年，砷化镓肖特基栅场效应晶体管(GaAs MESFET)问世。其后，随着异质结技术和超薄外延生长技术的日趋成熟，相继出现了机理各异、结构新颖、性能优异的各种器件。除GaAs MESFET有了长足进步之外，又接连开发了异质结双极晶体管(HBT)和高电子迁移率晶体管(HEMT)，这些器件已经成功地用于毫米波及超高速电路。已经或正在开发的还有新一代电子器件——量子电子学器件，包括共振隧穿器件和各种介观器件。微电子学发展的现状和趋势说明，这一领域生机勃勃，预期在本世纪上半叶器件的理论和制造技术仍将不断取得突破性进展。

光电子学是在电子学的基础上吸收了光技术而形成的一门高技术学科，主体是半导体光电子学。半导体激光器在1962年发明并在1970年实现室温下连续工作，以及后来各种半导体光电器件的出现，逐渐形成了以半导体激光器和探测器为主体的光电子学。进入20世纪80年代以后，随着光通信等应用技术的迅猛发展，半导体光电产品的市场迅速扩大，光盘、激光打印机、光通信系统和家用摄像机等都是家喻户晓的产品。普遍预测，到2010年，光电子产品市场将达到可以与集成电路市场相比的程度。由于光电子信息技术和产业具有强大的生命力，半导体光电子器件的研究和开发将更加受到重视。随着原子级薄膜生长和纳米级微细加工技术的发展，通过人工设计的材料如异质、超晶格及量子阱(线、点)材料来研制新器件，将是半导体光电子学器件今后的发展趋势。

20世纪90年代将近10年的时间内，我曾在北京大学计算机科学技术系为微电子学专业的三年级本科生和一年级硕士生讲授“半导体器件物理”课程，在上述认识的基础上，组织了这门课程的教学，最后编写了这本书。

本书远非包揽所有半导体器件。半导体器件的种类很多，广泛应用于社会、经济和军事的

各个领域。本书只是以电子信息技术为主干线，选择一些用途最广和最具代表性的器件，讲述理解器件行为所需的一些基础知识，因此取名为“半导体器件物理基础”。在编写时，我特别注意物理图像清晰，尽力做到叙述上正确，既简明扼要地阐明学习器件必需的半导体知识及器件的基本特性，也用尽可能简化的模型讨论器件发展中新的物理效应和介绍新器件，力图使学生不仅能获得充足的基础知识，而且建立一些新的观念，能够适应半导体器件的现代发展。

本书应教材建设需要编写，得到了王阳元院士、韩汝琦教授和赵宝瑛教授的鼓励与关心。赵宝瑛、甘学温教授细致地审阅了全书，傅春寅、吴恩教授也阅读了部分章节，他们的热情帮助使作者受益颇多。北京大学出版社沈承凤老师和其他有关工作人员为本书的出版付出了辛勤劳动。在此，作者一并致谢。

对我来说，写这样一本书的确是一项艰巨的工作，为之耗费了难以置信的时间。我也深感学识浅薄，本书肯定有许多不足、不妥和错误，诚恳希望有关方面的专家和读者指正。

作　者

2001 年 9 月

主要符号表

a	晶格常数
A	面积
A^*	里查孙常数
c	真空中的光速
C	电容;声速
d	厚度
D	扩散系数
D_n	电子扩散系数
D_p	空穴扩散系数
E	能量
E_c	导带底的能量
E_F	费米能级
E_{Fn}	电子准费米能级
E_{Fp}	空穴准费米能级
E_i	本征费未能级
E_g	禁带宽度(禁带隙)
E_v	价带顶的能量
$\mathscr{E}, \boldsymbol{\mathscr{E}}$	电场
E_c	临界电场,最大击穿电场
$\mathscr{E}_m$	最大电场
f	频率
$f(E)$	费米·狄拉克分布函数;占有几率
$f_n(E)$	电子占有几率
$f_p(E)$	空穴占有几率
g	态密度;电导
G	产生率;增益系数
h	普朗克常数
I	电流强度
I_C	集电极电流
I_{ph}	光电流
J	电流密度
J_{th}	阈值电流密度
J_n	电子电流密度
J_p	空穴电波密度
k_B	玻耳兹曼常量
$k, \boldsymbol{k}$	电子波矢
l	长度
L	扩散长度;长度
L_D	德拜长度
L_n	电子扩散长度
L_p	空穴扩散长度
m	有效质量
m_d	态密度有效质量
m_l	纵向有效质量
m_n	电子有效质量
m_p	空穴有效质量
m_{ph}	重空穴有效质量
m_{pl}	轻空穴有效质量
m_t	横向有效质量
m_0	自由电子质量
m_σ	电导有效质量
n	电子浓度
$\bar{n}$	实折射率
$\tilde{n}$	复折射率
n_i	本征载流子浓度
n_q	波矢为 q 的声子数
n_s	表面电子浓度
N	杂质浓度
N_A	受主杂质浓度
N_c	导带有效态密度
N_D	施主杂质浓度
N_v	价带有效态密度
p	空穴浓度
P	动量;散射几率
q	电子电荷;
$\boldsymbol{q}$	声子波矢

Q_b	耗尽层电荷面密度
Q_g	栅电荷面密度
Q_i	反型层电荷面密度
Q_0	有效界面电荷
Q_s	半导体表面层电荷面密度
R	电阻;反射率
t	时间;厚度
T	绝对温度
v, $\boldsymbol{v}$	速度
v_c	热发射速度
v_n	电子漂移速度
v_p	空穴漂移速度
v_s	饱和速度
v_{th}	热运动速度
V	电压
V_{bi}	内建电势
V_{BE}	基极-发射极电压
V_{BC}	基极-集电极电压
V_B	击穿电压
V_{FB}	平带电压
V_T	阈值电压
W	中性基区宽度
α	吸收系数;共基极电流增益
β	其发射极电流增益
γ	发射效率;衬偏系数
δ	间隙宽度
Γ	光限制因子
ε_0	真空介电常数
ε_{ox}	氧化物介电常数
ε_s	半导体介电常数
τ	寿命或弛豫时间
η	转换效率;量子效率
θ	角度;渡越角
λ	波长;平均自由程
ν	光的频率
μ_n	电子迁移率
μ_p	空穴迁移率
ρ	电阻率;电荷密度
σ	电导率;俘获截面
σ_G	微分增益系数
ϕ_F	费米势
ϕ_{Bn}	n 型半导体的肖特基势垒高度
ϕ_m	金属功函数
ϕ_s	半导体功函数
χ	电子亲和势
ψ	电势;电子波函数
ω	角频率

目　录

主要符号表 …… (1)

第一章　半导体基本知识 …… (1)

1.1　半导体中的载流子 …… (1)

1.1.1　半导体的能带 …… (1)

1.1.2　载流子平衡浓度 …… (7)

1.2　晶格振动 …… (12)

1.2.1　晶格振动和格波 …… (12)

1.2.2　声子 …… (14)

1.2.3　晶格散射 …… (16)

1.3　载流子输运现象 …… (17)

1.3.1　载流子漂移 …… (17)

1.3.2　载流子扩散 …… (22)

1.3.3　电流密度方程(漂移-扩散方程) …… (23)

1.3.4　过剩载流子及其产生与复合 …… (25)

1.3.5　连续性方程和泊松方程 …… (30)

1.3.6　强电场效应 …… (31)

1.3.7　非稳态输运效应：速度过冲 …… (34)

1.4　半导体的光学性质 …… (35)

1.4.1　引言 …… (35)

1.4.2　辐射跃迁和光吸收 …… (36)

1.4.3　直接禁带半导体中带间跃迁的自发发射和载流子寿命 …… (38)

1.4.4　直接禁带半导体中的光吸收和光增益 …… (40)

第二章　p-n 结 …… (42)

2.1　热平衡状态 …… (42)

2.2　耗尽区和耗尽层电容 …… (45)

2.2.1　热平衡情形 …… (45)

2.2.2　非平衡情形 …… (48)

2.2.3　耗尽层电容(势垒电容) …… (49)

2.3　直流特性 …… (50)

2.3.1　理想伏-安特性 …… (50)

2.3.2　产生-复合效应 …… (52)

2.3.3　大注入效应 …… (53)

2.3.4　温度效应 …… (54)

2.4　交流小信号特性；扩散电容 …… (55)

2.5 电荷存储和反向恢复时间 …… (57)
2.5.1 存储电荷 …… (57)
2.5.2 瞬态特性和反向恢复时间 …… (58)
2.6 结的击穿 …… (59)
2.6.1 隧道击穿 …… (59)
2.6.2 雪崩击穿 …… (60)
第三章 双极型晶体管 …… (66)
3.1 基本原理 …… (66)
3.1.1 基本结构 …… (66)
3.1.2 放大工作状态 …… (66)
3.1.3 电流增益 …… (67)
3.2 双极型晶体管的直流特性 …… (69)
3.2.1 理想晶体管的电流 …… (69)
3.2.2 电流基本方程 …… (70)
3.2.3 放大状态 …… (71)
3.2.4 非理想现象分析 …… (73)
3.2.5 输出特性 …… (79)
3.3 双极型晶体管模型 …… (80)
3.3.1 E-M 模型 …… (80)
3.3.2 G-P 模型 …… (82)
3.4 双极型晶体管的频率特性 …… (85)
3.4.1 小信号交流等效电路 …… (85)
3.4.2 电流增益随频率的变化 …… (86)
3.4.3 频率参数 …… (87)
3.4.4 基区渡越时间 …… (88)
3.4.5 功率-频率限制 …… (89)
3.5 双极型晶体管的开关特性 …… (90)
3.5.1 晶体管的开关作用 …… (90)
3.5.2 关断和导通阻抗 …… (90)
3.5.3 导通时间和存储延迟时间 …… (91)
3.6 异质结双极晶体管(HBT) …… (94)
3.6.1 异质结的能带图 …… (94)
3.6.2 HBT 电流放大的基本理论 …… (96)
3.6.3 几类常见的 HBT …… (98)
3.7 多晶硅发射极晶体管(PET) …… (99)
3.7.1 能带图和物理参数 …… (100)
3.7.2 少子分布和电流密度 …… (101)
3.7.3 注入发射区的空穴电流密度 J_{P1} …… (102)
3.7.4 电流增益 …… (102)

3.8 p-n-p-n 结构 …… (103)
第四章 化合物半导体场效应晶体管 …… (106)
4.1 肖特基势垒和欧姆接触 …… (106)
4.1.1 肖特基势垒 …… (106)
4.1.2 肖特基势垒二极管 …… (108)
4.1.3 欧姆接触 …… (111)
4.2 GaAs MESFET …… (111)
4.2.1 基本结构 …… (111)
4.2.2 夹断电压和阈值电压 …… (112)
4.2.3 电流-电压特性 …… (113)
4.2.4 截止频率 …… (116)
4.3 高电子迁移率晶体管(HEMT) …… (117)
4.3.1 器件结构及特点 …… (117)
4.3.2 2DEG 浓度(面密度) …… (118)
4.3.3 HEMT 的基本性质 …… (120)
第五章 MOS 器件 …… (124)
5.1 MOS 结构的基本性质 …… (124)
5.1.1 基本结构和能带图 …… (124)
5.1.2 平带电压 …… (125)
5.1.3 表面势 …… (126)
5.1.4 电势平衡和电荷平衡 …… (126)
5.1.5 半导体表面状态 …… (127)
5.1.6 表面层电荷和表面势的关系 …… (128)
5.1.7 表面势和栅压的关系 …… (129)
5.1.8 表面反型状态 …… (129)
5.1.9 电容和电压的关系(C-V 特性) …… (130)
5.2 MOS 场效应晶体管的基本理论 …… (132)
5.2.1 基本结构和工作原理 …… (132)
5.2.2 非平衡状态 …… (134)
5.2.3 阈值电压 …… (135)
5.2.4 电流基本特性 …… (137)
5.2.5 瞬态特性 …… (143)
5.2.6 短沟道效应 …… (147)
5.2.7 CMOS(互补型 MOSFET)结构 …… (162)
5.3 短沟道 MOSFET …… (164)
5.3.1 器件小型化的规则 …… (164)
5.3.2 结深 X_j …… (166)
5.3.3 栅氧化层厚度 d_{ox} …… (167)
5.3.4 耗尽层宽度 $W_{S,D}$ …… (168)

5.3.5 最短沟道长度 L_{min} …… (170)
5.4 SOI MOSFET …… (171)
5.4.1 概述 …… (171)
5.4.2 阈值电压 …… (172)
5.4.3 电流特性 …… (173)
5.4.4 亚阈区斜率(摆幅) …… (174)
5.4.5 热电子效应 …… (175)
5.5 埋沟 MOSFET …… (176)
5.6 电荷耦合器件(CCD) …… (178)
5.6.1 表面深耗尽状态和电荷存储 …… (178)
5.6.2 基本的 CCD 结构 …… (179)
5.6.3 转移效率和频率响应 …… (181)
第六章 有源微波二极管 …… (183)
6.1 隧道二极管 …… (183)
6.1.1 隧道输运过程 …… (183)
6.1.2 隧道二极管 …… (188)
6.2 共振隧穿二极管(RTD) …… (191)
6.2.1 共振隧穿效应 …… (191)
6.2.2 电流-电压特性 …… (192)
6.2.3 微波性能 …… (194)
6.3 IMPATT 二极管 …… (194)
6.3.1 里德二极管 …… (194)
6.3.2 雪崩延迟和渡越时间效应(小信号分析) …… (196)
6.3.3 功率和效率(大信号分析) …… (198)
6.4 转移电子器件 …… (201)
6.4.1 电子转移和负微分迁移率 …… (201)
6.4.2 偶极畴和基本工作原理 …… (203)
6.4.3 器件工作状态 …… (207)
第七章 半导体激光器和光电二极管 …… (208)
7.1 半导体激光器的基本结构和工作原理 …… (208)
7.1.1 基本结构(光学谐振腔) …… (208)
7.1.2 增益、损耗和光振荡条件 …… (209)
7.2 半导体激光器的工作特性 …… (212)
7.2.1 阈值电流密度 …… (212)
7.2.2 输出功率和效率 …… (213)
7.2.3 调制特性 …… (214)
7.2.4 激光线宽 …… (218)
7.3 双异质结(DH)激光器 …… (220)
7.3.1 能带图和光场分布 …… (220)

7.3.2 半导体材料 …… (221)
7.3.3 改进的器件结构 …… (222)
7.4 量子阱(QW)激光器 …… (224)
7.4.1 量子阱中的载流子 …… (224)
7.4.2 量子阱中的光增益 …… (226)
7.4.3 应变层量子阱 …… (228)
7.5 光电二极管 …… (230)
7.5.1 p-n 结的光电流 …… (230)
7.5.2 p-i-n 光电二极管和雪崩光电二极管 …… (234)
7.5.3 光电二极管的特性参数 …… (236)
主要参考文献 …… (241)

第一章　半导体基本知识

作为理解各类半导体器件性质的基础，本章概述半导体的能带、载流子浓度、输运现象和光学性质。在标准的半导体物理学或固体物理学的教科书中，这些课题都得到了详细的论述，我们这里只是以比较直观的论证方式介绍一些基本概念和重要结论。

1.1　半导体中的载流子

1.1.1　半导体的能带

1. 能带　电子和空穴

按照固体的量子理论，当原子凝聚成为固体（并不只限于晶体）时，由于原子间的相互作用，相应于孤立原子的每个能级加宽成由间隔极小（准连续）的分立能级所组成的能带，能带之间隔着宽的禁带。图 1.1 为金刚石结构晶体如何从孤立的硅原子能级形成能带的示意图，图右端画出了孤立原子的两个分立能级（如果将这些孤立原子看成一个体系，那么每个分立能级都是简并的），随着原子间距的减小，各个简并的能级分裂，形成能带。当原子间距进一步缩小时，由各个不同的分立能级所形成的各能带失去其特性，合并成一个能带。当原子间距接近金刚石晶体中原子间的平衡距离（硅的晶格常数为 5.43Å）时，这一能带再次分裂成为两个能带。两个能带之间的区域表示固体中的电子不能具有的能量，所以称之为禁带或带隙，通常以 E_g 表示。在禁带上面的能带叫导带，在禁带下面的能带叫价带，如图 1.1 左端所示。

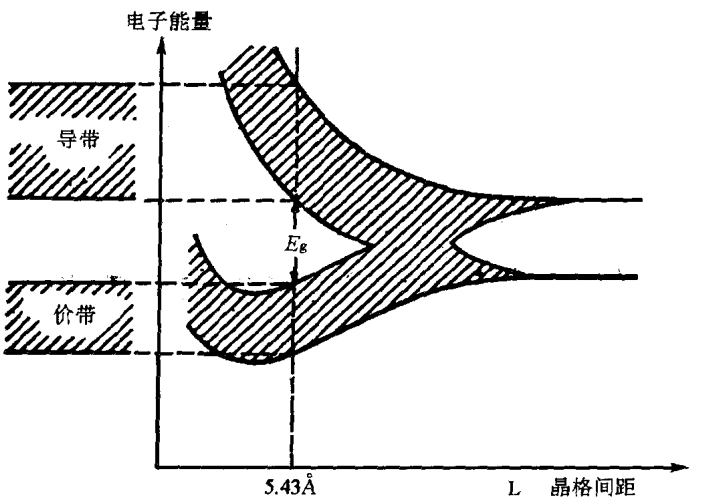


图 1.1　使孤立的硅原子彼此接近组成金刚石结构晶体时形成能带

固体的电学性质由原子最外围电子（价电子）填满或只是部分填充固体能带来决定。对于我们所讨论器件涉及的半导体，价电子实际上完全填满价带。因为能带填满了，就电子的整体

而言是不能获得能量的，所以不能产生净电流流动。因此，除非能够使一些电子离开价带，让位给少量电子以获得能量，否则是不可能导电的。另一方面，少量电子进入邻近的较高能带(导带)，导带中的电子将自由地获得能量，所以能够导电。

虽然在较高的温度下，热振动会使结合强度一般的半导体价键(不像绝缘体的价电子与最近邻原子形成强键)中的一些破裂，有少量价带电子越过禁带进入导带，但是这种本征过程是不重要的。用于器件的半导体材料主要是非本征的，即载流子是由掺入半导体的特定类型的杂质原子所产生的。如果载流子是导带中的电子，则杂质被称作施主，相应的半导体材料是 n 型；反之，如果半导体中含有收留来自价带电子(从而留给价带一个电子空位或者说一个空穴)的杂质，则杂质被称作受主，相应的半导体材料是 P 型。值得注意的是，应把空穴看作与电子类似，是具有动量和能量的量子波包(准经典粒子)。空穴带正电，并在外电场作用下运动，运动的方向与电子相反。形象地讲，空穴的概念和流体中的气泡类似，虽然实际上是液体在流动，但我们可以把它看成是气泡沿相反方向移动。

图 1.2 表示电子能量和空穴能量的关系。E_c 代表导带底的能量，相应于电子的势能，即导带电子静止时的能量；E_c 以上的能量表示电子的动能。E_v 代表价带顶的能量，相应于空穴的势能，E_v 以下的能量表示空穴的动能。E_c 和 E_v 之差即为禁带宽度：E_g 是半导体物理中最重要的参数。通常，能带图表示的是电子的能量。当电子能量增加时，电子跃迁到能带图上较高的位置；相反，空穴的能量增加时，空穴在价带内向下跃迁(这是因为空穴带正电，与电子相反)。

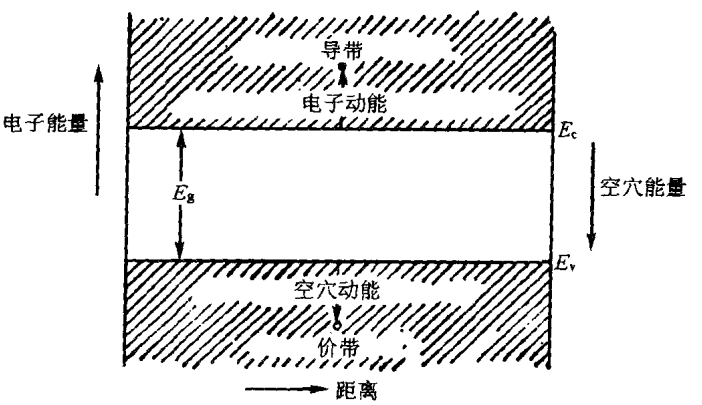


图 1.2　能带图中的势能与动能表示

实验结果表明，大多数半导体的禁带宽度随温度的升高而减小，禁带宽度随温度的变化可近似表示为

$$E_g(T) = E_g(0) - \alpha T^2/(T + \beta) \tag{1.1}$$

对 Si, Ge 和 GaAs，参数 $E_g(0)$，α 和 β 之值如表 1.1 所示。

表 1.1　$E_g(T)$表示式(1.1)中的参数值

半导体材料	$E_g(0)$(eV)	$\alpha(\times 10^{-4}\text{eV/K}^2)$	β(K)
Si	1.17	4.73	636
Ge	0.74	4.77	235
GaAs	1.52	5.41	204

在室温及常压下，硅的禁带宽度为 1.12eV，锗为 0.68eV，砷化镓为 1.42eV。

2. 有效质量近似

量子力学分析表明，晶体中电子的波函数为布洛赫函数，形式为

$$\psi_k(r) = u_k(r)e^{ik\cdot r} \tag{1.2}$$

k 是波矢，r 是空间矢量，$u_k(r)$是具有和晶格相同周期的周期函数。$u_k(r)$反映周期场对电子运动的影响。对于自由电子(动量 $P = \hbar k$，$\hbar = h/2\pi$ 是约化普朗克常数)，波函数退化为平面波 $\psi_k(r) = Ae^{ik\cdot r}$，$A$ 是与位置无关的常数。显然，与平面波不同，布洛赫函数并不对应于确定的动量，即晶体中波矢 k 的状态并不对应确定的动量，$P = \hbar k$ 不再具有严格意义下的动量的含义，而称之为晶体动量。

无疑，电子在晶体中的能级(本征值)和相应的布洛赫函数(本征态)是讨论各种有关电子问题的基础。但是，可以证明，对于一般的输运过程问题(例如电、磁场中各种电导效应)，可以把电子(或空穴)看成具有动量 P(对于 k 的状态，$P = \hbar k$)和能量 $E = P^2/2m$ 的粒子(量子波包)，m 称为有效质量。这种方法称为有效质量近似。

E-k 关系

已经采用各种数值方法对固体的能带(E-k 关系)进行了理论研究。图 1.3 示出了 Ge, Si 和 GaAs 的能带图。下面只是从有效质量近似出发，对 E-k 关系进行一些讨论。

在有效质量近似下，电子的动量 P 和能量 E 分别为

$$P = \hbar k \tag{1.3}$$

$$E = (\hbar^2/2m)k \cdot k \tag{1.4}$$

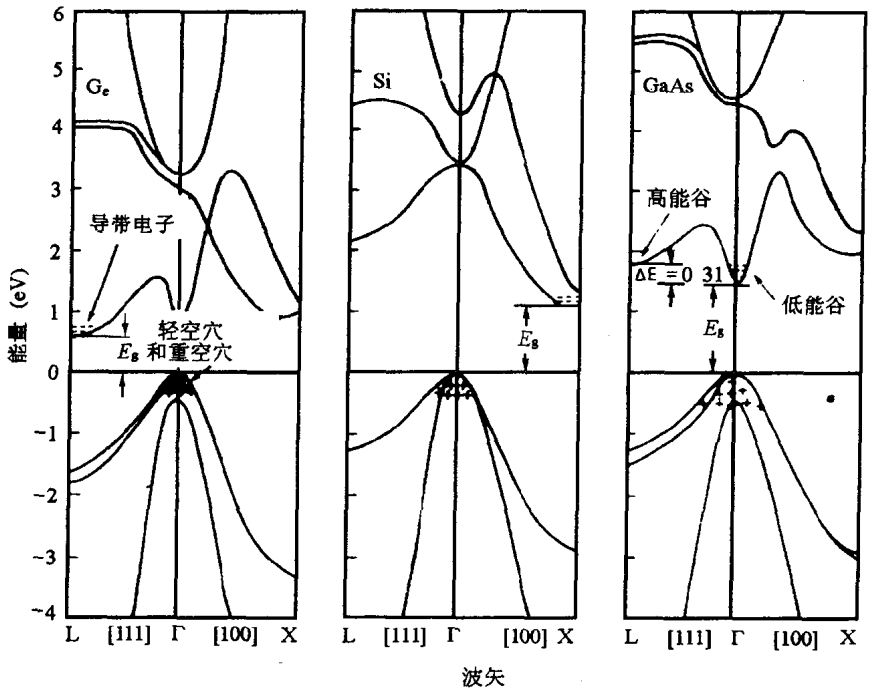


图 1.3　Ge, Si 和 GaAs 的能带结构

一般，$E(k)$在 $k = k_0$ 处有极值(极大值或极小值)E_0。在导带中，电子倾向在最低能量

处，因此其行为可以用在最小能量附近的级数展开式来描述：

$$E - E_0 = (\hbar^2/2m)(k - k)^2 \tag{1.5}$$

如果画出能量 $E(k)$ = 常数的三维表面，它将是一个圆球，即所谓球形等能面。画等能面的空间叫做 k 空间。

在更普遍的情形下，k 空间的等能面是椭球面，即方程(1.5)应修改成

$$E - E_0 = (\hbar^2/2m_1)(k_1 - k_{10})^2 + (\hbar^2/2m_2)(k_2 - k_{20})^2 + (\hbar^2/2m_3)(k_3 - k_{30})^2 \tag{1.6}$$

这里必须适当地选择各轴在 k 空间的方向。普遍说来，有效质量是一个张量，m_1, m_2, m_3 是主质量。

热电子是能量远远超过晶体中电子的正常平均热能的电子，其能量并不局限在最小能量的区域中，有效质量近似显然有很大的误差。但是，由于这种近似用起来很方便，并且已经证实甚至在理解高能电子现象时也有实用价值，所以通常使用这一近似方法。

空穴实际上是价带中最高能量上面的电子波波包，等效于式(1.5)的展开式为

$$E - E_0 = -(\hbar^2/2m)(k - k_0)^2 \tag{1.7}$$

上式中的负号表示最大值。这些电子具有负有效质量。然而，如果全部带负电荷和负有效质量的电子被电场扫动的话，则波包的运动与一个带正荷和正有效质量的粒子的运动是不可区分的。这个粒子就是空穴。

布里渊区

前面的论述说明了半导体中的电子只能取一定允许的能量或动量值。如果我们考虑在一定方向上传播的平面电子波，在此方向上晶体中等效原子平面的间距为 a，则波函数在每一个周期的平面上都有反射，而当 a 等于半波长 $\lambda/2$ 时，被邻近原子反射的子波位相相差 π。但是当被 B 反射的子波到达被 A 反射的子波时，它们的位相相同。这点也适用于其他的子波。这样，所有反射的子波相长地干涉，如图 1.4 所示。结果形成电子波的全部反射，即当波矢量满足条件

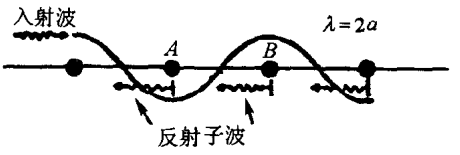


图 1.4　晶体中反射波的相长干涉

$$|k_m| = \frac{2\pi}{\lambda} = \frac{\pi}{a} \tag{1.8}$$

时，能量的传播成为不可能。因为晶体的周期性随方向而变，传播停止的 $|k_m|$ 值也随方向而变。在 k 空间中，由所有这些 k_m 矢量形成的空间叫做布里渊区(第一布里渊区，或简约布里渊区)。

图 1.5 画出了面心立方晶格(例如金刚石和闪锌矿晶格)的第一布里渊区。元素半导体 Si 和 Ge 具有金刚石结构，而大多数Ⅲ-Ⅴ族化合物半导体(例如 GaAs)具有闪锌矿结构晶格。图 1.5 也标出了布里渊区的对称点和对称线。例如，Γ 表示 $k = 0$(布里渊区中心)的点，L 表示沿〈111〉轴(Λ)的边缘点，X 表示沿〈100〉轴(Δ)的边缘点，K 表示沿〈110〉轴(Σ)的边缘点。

值得注意的是电子的动量 $\hbar k$ 可以超过式(1.8)定义的 $\hbar k_m$，但电子的能量是由 $k - n k_m$ 来计算的，其中整数 n 的选择条件是使得总的波矢量位于布里渊区内。换句话说，$E(k)$ 是随 k 周期性地重复的函数，周期性的单元由布里渊区确定。正因为如此，图 1.3 限定的 k 值足以描述电子的 E-k 关系。

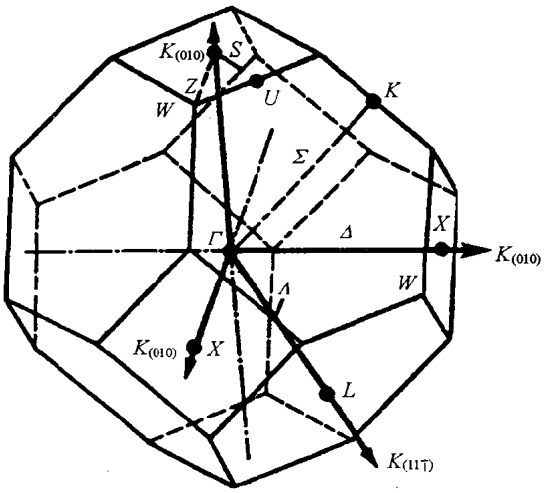


图 1.5　金刚石和闪锌矿晶格的布里渊区

多能谷半导体

许多重要的半导体(例如 Ge, Si, GaAs)不是只有一个导带极小值,而是有若干个位于 $\boldsymbol{k}$ 空间不同点的极小值。例如, Si 具有位于〈100〉轴上 $k=k_0$ 的 6 个极小值(k_0 约等于最大波矢量的 0.85 倍),相应地有 6 个能谷(极小值为能谷的谷底),如图 1.6 所示。通常, E-$\boldsymbol{k}$ 围绕极小值不是球对称,而是旋转椭球对称。

$$E-E_c=\frac{\hbar^2}{2m_l}(k_l-k_0)^2+\frac{\hbar^2k_t^2}{2m_t} \qquad (1.9)$$

E_c 为谷底(导带底)能量, k_l 是沿椭球长轴方向(纵向)的波矢, k_t 是横向波矢, m_l 和 m_t 分别是纵向和横向有效质量。

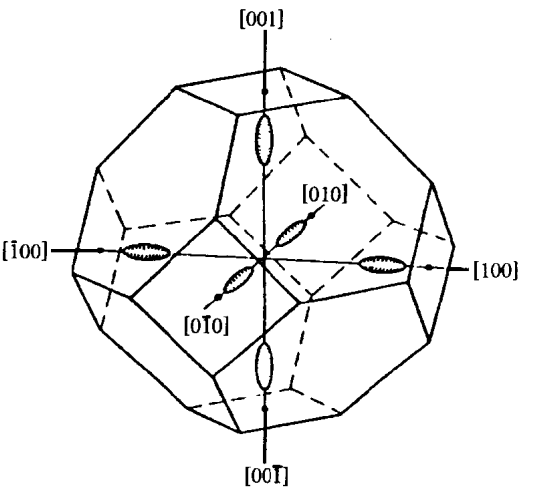


图 1.6　Si 中导带的椭球等能面

上述 Si 的导带能谷都具有同样的极小值和性质,平均来说能容纳同样数目的电子,所以称这些能谷是等价的。

在 GaAs 中,导带极小值在 Γ 点($\boldsymbol{k}=0$),等能面是球面,

$$E-E_c=\frac{\hbar^2}{2m_n}k^2 \qquad (1.10)$$

k 为波矢数值, m_n(约 $0.067m_0$, m_0 是自由电子质量)为电子有效质量。除 Γ 处外,沿〈111〉轴接近 L 点还有能量稍高(约高 0.31eV)但有效质量大得多的 8 个等价能谷。在强电场下获得足够高的能量时,电子可以由 Γ 能谷(低能谷)向 L 能谷(次能谷)转移,产生所谓电子转移效应。

对于价带,普遍的情形是极大值发生在 Γ 点,并有两个简并化的能谷,从而产生两类空穴,即轻空穴(有效质量 m_{pl})和重空穴(有效质量 m_{ph})。这种简并性质来源于自旋和轨道角

动量之间耦合的量子力学效应。这一效应在量子阱激光器中是重要的。

表 1.2　Ge, Si, GaAs 中电和子空穴的有效质量

半导体材料	电　子	空　穴
Ge	$1.64(m_l/m_0)$	$0.04(m_{pl}/m_0)$
	$0.082(m_t/m_0)$	$0.28(m_{ph}/m_0)$
Si	$0.98(m_l/m_0)$	$0.16(m_{pl}/m_0)$
	$0.19(m_t/m_0)$	$0.49(m_{ph}/m_0)$
GaAs	$0.067(m_n/m_0)$	$0.082(m_{pl}/m_0)$
		$0.45(m_{ph}/m_0)$

注：(m_0：自由电子质量)

3. 态密度

每当我们想知道能谷中的电子数量时，就要知道能谷的量子态密度。这里简单考虑一下导带能谷的情况，类似的结果也适用于价带空穴。

在有效质量近似中，具有动量 $\hbar\boldsymbol{k}$ 的电子代表中心位置为 $\boldsymbol{k}$ 的小范围 $\Delta\boldsymbol{k}$ 的布洛赫波组成的量子波包，波包占据的空间范围 Δx 和波包所要求的量子波矢量范围 $\Delta\boldsymbol{k}$ 之间满足下述关系：

$$\Delta x \cdot \Delta k = 2\pi \tag{1.11}$$

在三维情形下有：

$$V \cdot V_{\mathbf{k}} = (2\pi)^3 \tag{1.12}$$

由此得出占据空间体积 V 的一个量子态在 $\boldsymbol{k}$ 空间所要求的体积 $V_{\mathbf{k}}$。由于具有相同波矢量的电子具有不同的自旋，每个 $\boldsymbol{k}$ 矢量存在两种状态，所以式(1.12)应修改为

$$2V \cdot V_{\mathbf{k}} = (2\pi)^3 \tag{1.13}$$

现在，对于单位实际空间体积($V=1$)，我们来计算能量在 E 和 $E+\mathrm{d}E$ 之间的量子态数目 $g(E)\mathrm{d}E$。对位于 $\boldsymbol{k}_0$ 的球对称能谷，由式(1.5)决定的在 $\boldsymbol{k}$ 空间中球的半径为

$$\left(\frac{2m_n}{\hbar^2}\right)^{1/2}(E-E_c)^{1/2} \tag{1.14}$$

相应的球的体积为

$$\frac{4\pi}{3}\left(\frac{2m_n}{\hbar^2}\right)^{3/2}(E-E_c)^{3/2} \tag{1.15}$$

球内的量子态数($V=1$)为

$$\frac{2}{(2\pi)^3}\cdot\frac{4\pi}{3}\left(\frac{2m_n}{\hbar^2}\right)^{3/2}(E-E_c)^{3/2} \tag{1.16}$$

当 E 增加时得到能量在 E 和 $E+\mathrm{d}E$ 间隔内的量子态总数，即

$$g(E)\mathrm{d}E = \frac{1}{2\pi^2}\left(\frac{2m_n}{\hbar^2}\right)^{3/2}(E-E_c)^{1/2}\mathrm{d}E \tag{1.17}$$

其中能态密度分布函数为

$$g_c(E) = \frac{1}{2\pi^2}\left(\frac{2m_n}{\hbar^2}\right)^{3/2}(E-E_c)^{1/2} \tag{1.18}$$

E_c 为谷底(导带底)的能量。

对椭球对称的能谷，$E-k$ 关系已由方程(1.6)给出，等能面是以(k_{10}, k_{20}, k_{30})为中心的椭球面。能量为 E 的椭球面的方程为

$$\frac{(k_1-k_{10})^2}{\frac{2m_1}{\hbar^2}(E-E_c)}+\frac{(k_2-k_{20})^2}{\frac{2m_2}{\hbar^2}(E-E_c)}+\frac{(k_3-k_{30})^2}{\frac{2m_3}{\hbar^2}(E-E_c)}=1 \tag{1.19}$$

各项的分母等于椭球各个半轴长的平方，因此椭球的体积为

$$\frac{4\pi}{3}\left(\frac{8m_1m_2m_3}{\hbar^2}\right)^{1/2}(E-E_c)^{3/2} \tag{1.20}$$

我们注意到，如果在式(1.20)中令$(m_1m_2m_3)^{1/3}=m_n$，则就是式(1.15)。所以不难理解，如果引入

$$m_{dn}=(m_1m_2m_3)^{1/3} \tag{1.21}$$

代替式(1.18)中的 m_n，就可以得到椭球等能面能谷(球形等能面能谷是它的特殊情形)的能态密度分布函数：

$$g_c(E)=\frac{1}{2\pi^2}\left(\frac{2m_{dn}}{\hbar^2}\right)^{3/2}(E-E_c)^{1/2} \tag{1.22}$$

m_{dn}称为电子的态密度有效质量。如果存在 s 个等价能谷，状态密度应该包括所有这 s 个能谷的贡献。把这种效应算进态密度有效质量，则有

$$m_{dn}=(s^2m_1m_2m_3)^{1/3} \tag{1.23}$$

GaAs 只有 $\Gamma(k=0)$处一个导带能谷，等能面是球面，$m_{dn}=m_n$；Si 和 Ge 都有多个导带能谷，等能面是旋转椭球面，$m_{dn}=(s^2m_lm_t^2)^{1/3}$，$m_l$ 和 m_t 分别为纵向和横向有效质量。

1.1.2 载流子平衡浓度

半导体中的电子和空穴通常称为载流子，因为输运电流的正是这些粒子。载流子的浓度决定了半导体的电导率，是半导体的一个重要参量。为了确定载流子浓度，需要利用统计物理的基本结果。

1. 费米分布函数

根据统计物理，能级为 E 的电子态能被电子占据的几率由费米分布函数给出：

$$f(E)=\frac{1}{1+e^{(E-E_F)/k_BT}} \tag{1.24}$$

式中，k_B 为玻耳兹曼常量，T 为绝对温度，E_F 为费米能级。E_F 的物理意义是，该能级上的一个状态被电子占据的几率为 1/2。当能量比 E_F 大于 $3k_BT$ 时，式(1.24)可简化为

$$f(E)\approx e^{-(E-E_F)/k_BT} \tag{1.25}$$

这正是经典的玻耳兹曼分布。

2. 载流子浓度

我们用下述方法来计算导带中的电子浓度：在能量 E 至 $E+dE$ 范围内的状态数为 $g_c(E)dE$，这里的 $g_c(E)$是能态密度，由式(1.22)给出；因为每个态被电子占据的几率为 $f(E)$，所以出现在此能量范围内的电子浓度等于 $f(E)g_c(E)dE$，对整个导带积分便得到导带中的电子浓度 n，即

$$n = \int_{E_c} f(E) g_c(E) \mathrm{d}E \tag{1.26}$$

将式(1.24)和(1.22)代入，并考虑到高能量时被积函数按指数下降，从而把积分上限取作无限大而不致引起明显误差，可得

$$\begin{aligned} n &= \frac{1}{2\pi^2}\left(\frac{2m_{dn}}{\hbar^2}\right)^{3/2}\int_{E_c}^{\infty}\frac{(E-E_c)^{1/2}\mathrm{d}E}{1+\mathrm{e}^{(E-E_F)/k_BT}} \\ &= N_c F_{1/2}(\eta_c) \end{aligned} \tag{1.27}$$

式中，N_c 称为导带有效态密度，

$$N_c = 2\left(\frac{m_{dn}k_BT}{2\pi\hbar^2}\right)^{3/2} \tag{1.28}$$

$F_{1/2}(\eta_c)$为 1/2 阶费米积分，

$$F_{1/2}(\eta_c) = \left(\frac{2}{\sqrt{\pi}}\right)\int_0^{\infty}\frac{\xi^{1/2}\mathrm{d}\xi}{1+\mathrm{e}^{\xi-\eta_c}} \tag{1.29}$$

$$\xi = \frac{E-E_c}{k_BT},\quad \eta_c = \frac{E_F-E_c}{k_BT} \tag{1.30}$$

$F_{1/2}(\eta)$有下述性质：当 $\eta \ll -1$ 时，有

$$F_{1/2}(\eta) \approx \mathrm{e}^{\eta} \tag{1.31}$$

当 $\eta \gg 1$ 时，有

$$F_{1/2}(\eta) \approx \frac{4\eta^{3/2}}{3\sqrt{\pi}} \tag{1.32}$$

以后我们将利用 $F_{1/2}(\eta)$的这种性质，讨论非简并和简并两种极限情形下的载流子浓度。

上述对导带中电子浓度的考虑完全适用于价带空穴。因为价带中的能量 E 的状态被电子占据的几率是$f(E)$，所以它被一个空穴占据的几率 $f_p(E)$为

$$\begin{aligned} f_p(E) &= 1 - f(E) \\ &= \frac{1}{1+\mathrm{e}^{(E_F-E)/k_BT}} \end{aligned} \tag{1.33}$$

空穴的能态密度 $g_v(E)$可以按照类似于推导电子能态密度 $g_c(E)$的过程得到，为

$$g_v(E) = \frac{1}{2\pi^2}\left(\frac{2m_{dp}}{\hbar^2}\right)^{3/2}(E_v - E)^{1/2} \tag{1.34}$$

m_{dp}为空穴的态密度有效质量，考虑到空穴总数应包括轻、重空穴带，有

$$m_{dp} = ({m_{pl}}^{3/2} + {m_{ph}}^{3/2})^{2/3} \tag{1.35}$$

m_{pl}, m_{ph}分别为轻、重空穴有效质量。还应注意，在式(1.34)中 $E < E_v$(E_v 为价带顶的能量)。空穴浓度为

$$\begin{aligned} p &= \int f_p(E) g_v(E) \mathrm{d}E \\ &= \frac{1}{2\pi^2}\left(\frac{2m_{dp}}{\hbar^2}\right)^{3/2}\int\frac{(E_v-E)^{1/2}\mathrm{d}E}{1+\mathrm{e}^{(E_F-E)/k_BT}} \end{aligned} \tag{1.36}$$

积分遍及整个价带，即从价带的最低能量积至价带顶。如果令

$$\xi = \frac{E_F-E}{k_BT},\quad \eta_v = \frac{E_v-E_F}{k_BT} \tag{1.37}$$

则式(1.36)可写作

$$P = \frac{1}{2\pi^2}\left(\frac{2m_{dp}k_BT}{\hbar^2}\right)^{3/2}\int_0^\infty \frac{\xi^{1/2}d\xi}{1+e^{\xi-\eta_v}}$$
$$= N_vF_{1/2}(\eta_v) \tag{1.38}$$

N_v 称为价带有效态密度,

$$N_v = 2\left(\frac{m_{dp}k_BT}{2\pi\hbar^2}\right)^{3/2} \tag{1.39}$$

3. 简并和非简并半导体

当费米能级 E_F 位于禁带内并且至能带边的距离大于若干个 k_BT($\eta_c < -3$, $\eta_v < -3$)时,半导体被称作是非简并的。这时有 $F_{1/2}(\eta) \approx e^\eta$,故

$$n = N_ce^{-(E_c-E_F)/k_BT} \tag{1.40}$$

$$p = N_ve^{-(E_F-E_v)/k_BT} \tag{1.41}$$

相反,当 E_F 进入导带或价带内时,半导体是简并的。这时有

$$F_{1/2}(\eta) \approx \frac{4}{3\sqrt{\pi}}\eta^{3/2}$$

若 E_F 进入导带内,则有

$$n = \frac{1}{3\pi^2}\left[\frac{2m_{dn}(E_F-E_c)}{\hbar^2}\right]^{3/2} \tag{1.42}$$

E_F 进入价带内,则有

$$p = \frac{1}{3\pi^2}\left[\frac{2m_{dp}(E_v-E_F)}{\hbar^2}\right]^{3/2} \tag{1.43}$$

对于非简并半导体,由式(1.40)和(1.41)相乘,得到

$$np = N_cN_ve^{-E_g/k_BT} \tag{1.44}$$

这一结果表明,乘积 np 和费米能级 E_F 无关,只决定于导带和价带的态密度、禁带宽度和温度。另一方面,式(1.44)也可以写作

$$np = n_i^2 \tag{1.45}$$

其中 n_i 称为本征载流子浓度,

$$n_i = (N_cN_v)^{1/2}e^{-E_g/2k_BT} \tag{1.46}$$

n_i 强烈依赖于温度 T,随温度上升而指数增加。室温时,硅的 n_i 为 $1.45\times10^{10}cm^{-3}$,砷化镓的 n_i 为 $1.79\times10^6cm^{-3}$。半导体的禁带宽度越大,本征载流子浓度越小。

4. 本征和非本征半导体

如果半导体十分纯净,以致电子和空穴的浓度相等,则称之为本征半导体,意思是载流子浓度决定于半导体的固有(本征)性质。反之,半导体掺入杂质以后就成为非本征半导体(掺杂半导体),在禁带内引入了杂质能级。

本征半导体 在本征半导体中,载流子只能通过将价带电子激发到导带来产生,即电子和空穴成对产生,从而浓度相等,为

$$n = p = n_i \tag{1.47}$$

因 n_i 随温度升高迅速增大,在足够高的温度下,一切半导体(除非杂浓度异常高)都将变成是

本征的。

n型半导体 以非本征的硅为例。当原来被 Si 原子(基质原子)占据的某些格点被五价杂质原子(例如 As 原子)占据时,As 的五个价电子中有四个参与同邻近的四个 Si 原子(Si 原子是四价的)形成共价键,第五个价电子不能进入已经饱和的键,它从杂质原子中分离出去,并像一个自由电子那样在整个晶体中运动,亦即此电子进入导带。由于硅中增加了带负电荷的载流子,所以成为 n 型半导体,As 原子称为施主。As 原子现在实际上是一个正离子 A_s^+,它有俘获一个自由电子的趋势,但由于它对电子的束缚能很小(像硅中的 As 这类施主称为浅施主,能级位置在导带底之下仅仅约 0.01eV),只要是温度不太低(不低于 100K)就不足以俘获电子。所以,如果温度高到一定程度,以致热能足以使浅施主电离、但本征激发的载流子浓度 n_i 还远低于掺杂浓度(室温就属于这种情形),则 n 型半导体(设浅施主浓度为 N_D)中的电子浓度为

$$n_n = N_D \tag{1.48}$$

下标 n 表示 n 型半导体,空穴浓度可由式(1.45)得到,为

$$p_n = n_i^2/N_D \tag{1.49}$$

P型半导体 仍以非本征硅为例。如果是三价原子(例如 B 原子)占据了硅中某些 Si 原子的格点位置,则 B 原子与邻近的 Si 原子形成共价键时,有一个键是空着的,此空键可能接受来自另一个键中的电子,于是后一个键处将出现一个空位(空穴)。此空穴在整个晶体中自由移动,亦即它进入价带。由于硅中增加了带正电荷的载流子(空穴),所以成为 P 型半导体,B 原子称为受主。受主在俘获一个附加电子后成为负离子,由于产生的空穴带有正电荷,它将受到受主的吸引作用。B 在硅中是浅受主,能级位于禁带中稍高于价带顶(约为 0.01eV)处,因此室温下几乎全部电离。当一个受主被电离(从价带顶激发一个电子填充其空键)时,便有一个空穴落入价带顶,成为一个自由载流子。按电子能量标度,此电离过程既可以说成是电子向上的跃迁,也可以说成是空穴向下的跃迁。与 n 型半导体类似,对 P 型半导体(设浅受主浓度为 N_A),室温时的空穴浓度为

$$p_p = N_A \tag{1.50}$$

电子浓度为

$$n_p = n_i^2/N_A \tag{1.51}$$

其中下标 P 表示 P 型半导体。

5. *费米能级*

由前面的讨论可知,费米能级 E_F 是确定半导体中载流子浓度的最重要的参数。对于本征半导体,$n = p$,令式(1.40)与式(1.41)相等,即得

$$E_F = E_i = \frac{E_c + E_v}{2} + \frac{3}{4}k_B T \ln(m_{dp}/s_{dn}) \tag{1.52}$$

室温时,上式右边第二项比禁带宽度小得多,因此本征半导体的费米能级(即本征费米能级)E_i 接近禁带中线。对于掺杂半导体,室温下 E_F 远离禁带中线,但多数情形(非简并情形)下处于禁带内某处,并且随半导体越来越变成 n 型时接近导带,越来越 P 型时移向价带。另一方面,由于本征载流子浓度随温度上升而指数增加,无论半导体是 n 型或 p 型的,温度高到一定程度以后将变成是本征的(此时本征载流子浓度远远超过施主浓度或受主浓度),即随着温度升高,对于一定的施主浓度或受主浓度,半导体的费米能级将称向禁带中线。图 1.7 对硅表

示了费米能级随杂质浓度和温度的变化关系，还标明了禁带宽度随温度的变化。

在讨论非本征半导体时，常用 E_i 作为参考能级。对于非简并情形，用本征载流子浓度 n_i 和本征费米能级 E_i，可将电子和空穴浓度表示为

$$\frac{n}{n_i} = e^{(E_F - E_i)/k_B T} \tag{1.53a}$$

和

$$\frac{p}{n_i} = e^{-(E_F - E_i)/k_B T} \tag{1.53b}$$

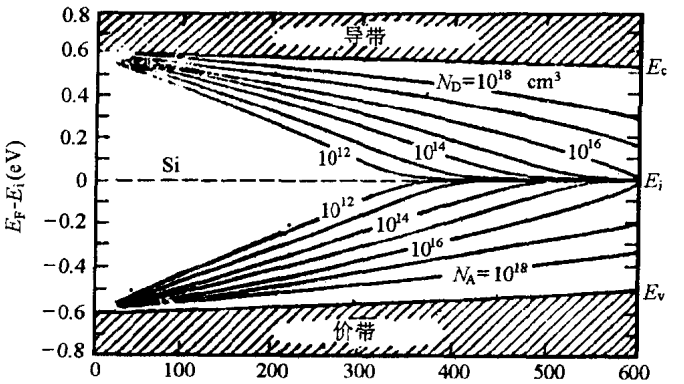


图 1.7　Si 的费米能级随杂质浓度和温度的变化关系

在掺杂浓度并非很低的一般情形下，玻耳兹曼分布并不是一个很好的近似，与之相应的表示式(1.40)和(1.41)是不精确的。在载流子浓度和费米能级之间，一个有用的表示式(Joyce-Dixon 近似)为

$$E_F = E_c + k_B T\left[\ln\frac{n}{N_c} + \frac{1}{\sqrt{8}}\frac{n}{N_c}\right] = E_v - k_B T\left[\ln\frac{p}{N_v} + \frac{1}{\sqrt{8}}\frac{p}{N_v}\right] \tag{1.54}$$

若载流子浓度 n 已知，上式可用来得到费米能级 E_F；反之，如果 E_F 已知，可用迭代法求解上式得到 n。当($n/\sqrt{8}N_c$)的项可以忽略时，结果对应玻耳兹曼近似。

6. 补偿半导体

实际上，所有半导体，甚至是很纯的半导体，都含有多种不同杂质，施主和受主二者并存。究竟是哪种导电类型，取决于过剩的是施主还是受主。这种施主和受主同时存在的半导体称为补偿半导体，因为一类掺杂剂的作用受到另一类掺杂剂“补偿”。最简单的补偿情形是半导体中只有一种施主杂质和一种受主杂质。在这种情形下，若 $N_A > N_D$，则有效受主浓度为

$$N_{Aeff} \approx N_A - N_D \tag{1.55}$$

反之，$N_D > N_A$，则有效施主浓度为

$$N_{Deff} \approx N_D > N_A \tag{1.56}$$

正如从式(1.55)或(1.56)所见，补偿作用可用来降低有效掺杂浓度，从而降低半导体的电导率。为明确起见，我们考虑 $N_D < N_A$，并且是深受主(能级位置靠近禁带中线的受主)。在这种情形下，来自浅施主杂质的电子全部掉进了受主能级，即施主被补偿了，材料的行为类似于本征半导体；尤其是宽禁带情形下更类似于绝缘体。例如 GaAs，以氧或铬作为深能级杂质

形成的所谓“半绝缘”GaAs 材料，就是依靠了这种补偿过程，其电阻高达 $10^8\Omega\cdot cm$。

乍一看来，施主浓度(N_D)和受主浓度(N_A)都很大且近乎相等的半导体很像本征半导体。实际上，杂质进入半导体材料是一个统计过程，施主浓度和受主浓度随位置局部起伏。当 N_D 和 N_A 很大时，这种杂质浓度起伏导致显著的电势起伏，能带图看起来会像一组在空间无规则分布的波峰和波谷，如图 1.8 所示。强补偿半导体有不少诱人的异常性质。例如，就像图 1.8 所表示的那样，费米能级在某些区域可能进入导带，而别的一些区域又可能在价带内。这意味着将形成电子和空穴的“液滴”，从而这种强补偿材料在低温下的行为像一个嵌入金属小颗粒的绝缘体。而且，在如此强补偿的材料中，低能量的导带电子受导带波峰阻挡而被定域，但高能导带电子可以穿越波峰。把这些低能和高能电子分开的临界能量称为渗透能级，它相对费米能级的位置决定强补偿半导体的电导率。

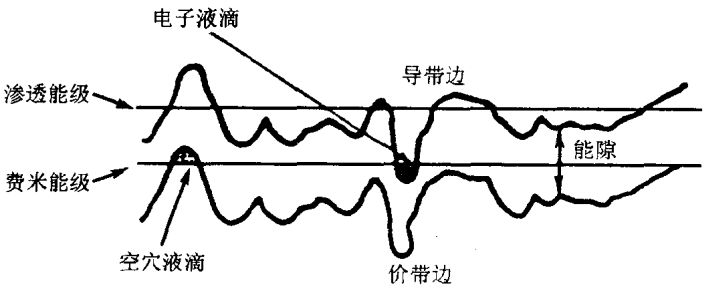


图 1.8　强补偿半导体中能带随位置的变化；电子势能起伏由施主及受主浓度的起伏引起

1.2　晶格振动

至此一直假定，半导体是完整晶体，原子处在严格的固定位置(格点)上。在这种理想的周期性晶格中，电子以有效质量自由运动。实际上，晶体原子离开其平衡位置的热振动，以及缺陷、杂质、位错和晶体边界等，都会影响电子的运动，使半导体产生电阻。其中，晶格热振动是半导体电阻的最基本的原因。

1.2.1　晶格振动和格波

晶体中原子的振动不是各自独立进行的，恰恰相反，它们可以分解成一系列在晶体中传播的波，称为晶体振动膜或格波。每个格波是以其波长 λ(或波数 $q=2\pi/\lambda$)和振动频率 ν(或 $\omega=2\pi\nu$)来表征的。在 Ge 和 Si 中，尤其明显地在 GaAs 中，每个元胞(晶体周期性的最小单元)都有二个不等价的原子，因此原子的周期点阵将组成所谓的双原子链。对于一维双原子链，振动频率为

$$\omega^2 \pm = \beta\left(\frac{1}{M_1}+\frac{1}{M_2}\right) \pm \beta\left[\left(\frac{1}{M_1}+\frac{1}{M_2}\right)^2 - \frac{4\sin^2(qa)}{M_1M_2}\right]^{1/2} \tag{1.57}$$

其中，a 为两个相邻原子的距离，β 为原子之间的力常数，其典型值约为 10^3 达因/厘米。与公式(1.57)的正、负号对应，图 1.9 画出各自不同值 q 的格波频率 ω_+，ω_-，属于 ω_+ 的格波分支

称为光学支，属于 ω_- 的格波分支称为声学支。

声学支和光学支之间的差异可以通过比较它们在长波极限（$q\to 0$）的性质清楚地看到。对于声学波，当（$q\to 0$）时，由式(1.57)可得

$$\omega_- \approx \left(\frac{2\beta}{M_1+M_2}\right)^{1/2} aq = Cq \qquad (1.58)$$

上式表明，长声学波的频率正比于波数，这正是弹性波的特点。声学波的群速（$v_g=\partial\omega/\partial q$）和相速（$v_p=\omega/q$）为

$$C = \left(\frac{2\beta}{M_1+M_2}\right)^{1/2} a \qquad (1.59)$$

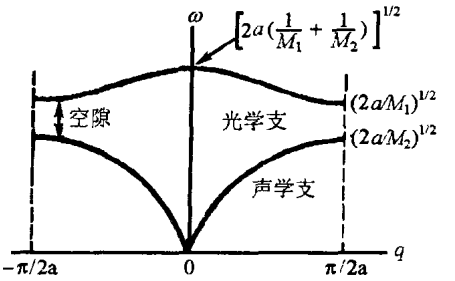


图 1.9　双原子晶格的光学波和声学波（$M_1<M_2$）

相应于 5Å 左右的晶格周期和约 10^{-23}g 的原子质量，C～5×10^3m/s，而声学波的最高频率近似为 5×10^{12}Hz。对于光学波，当 $q\to 0$，

$$\omega_0 \approx \left[2\beta\left(\frac{1}{M_1}+\frac{1}{M_2}\right)\right]^{1/2} \qquad (1.60)$$

其典型值 10^{13}Hz，这个频率在红外区内，相应的频率分支也就叫做光学支。

值得指出的是，在声学支中，元胞中的两种原子近似地作同位相运动，特别在长波极限下，两种原子的运动是完全一致的（振幅和位相均无差别）。而在光学支中，元胞中的两种原子基本上围绕质心作振动，特别在长波极限下，两种原子的振动有完全相反的位相而保持质心不动。图 1.10 形象地表示了长波极限下声学支和光学支中的原子位移。显然，如果元胞中只有一种原子，那就没有光学支。在 Ge，Si 和 GaAs 中，每个元胞中都有两个不等价的原子，所以都有光学支。

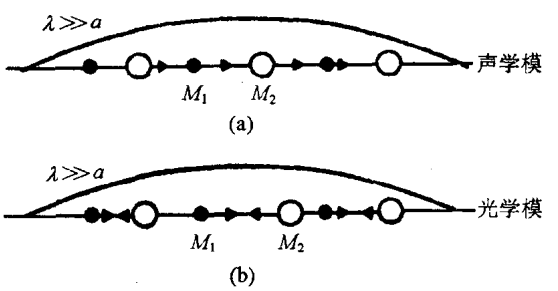


图 1.10　(a) 无限长波长（$q=0$）的声学模中的原子位移
(b) 无限长波长的光学模中的原子位移

还应当指出，在 $q=\pi/2a$ 处（这与前面讨论布里渊区的情形不同，实际点阵的周期是 $2a$，而不是 a）形成驻波，能量不能传播。这里，第一布里渊区是 $-\pi/2a\leqslant q\leqslant\pi/2a$。而由式(1.57)可知，在 q 空间中 $\omega(q)$ 是周期性的，周期为 π/a，即

$$\omega(q+\pi/a)=\omega(q) \qquad (1.61)$$

这说明相应于波矢 q 和 $q'=q+\pi/a$ 的两个格波表示相同的物理运动。为作出单值表示，并注意到 $\omega(-q)=\omega(q)$ 的反射对称性，将格波的波矢限制在刚才叙述的第一布里渊区内是十分方便的（参阅图 1.9）。

上面的讨论使用了一维振动模型，相应的格波只能是纵波。对于实际晶体，还应包括两个独立方向的横向振动。这样，对于元胞内包含两个原子的三维晶体，有一个纵声学支(LA)和一个纵光学支(LO)，它们来自每个原子沿波矢方向位移；有两个横声学支(TA)和两个横光学支(TO)，它们来自原子在垂直于波矢的平面内振动。图 1.11 示出了 Ge, Si 和 GaAs 的实测结果。

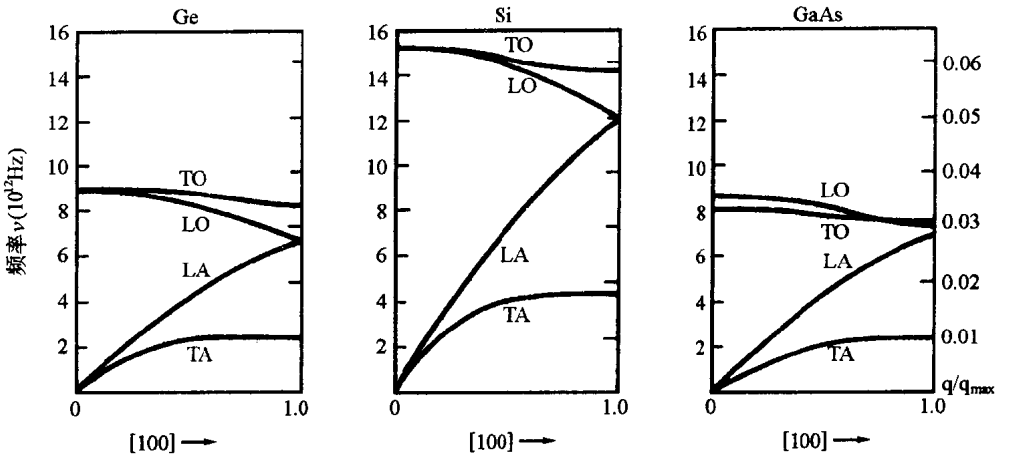


图 1.11　Ge, Si 和 GaAs 中的声子谱

1.2.2　声子

犹如光子是与电磁波相联系的量子粒子，声子是与格波相联系的量子粒子。与光子类似，对频率 ω 和波矢 $\boldsymbol{q}$ 的格波，声子的能量是 $\hbar\omega$，动量是 $\hbar\boldsymbol{q}$。因为每个格波的振动能量是量子化的，只能改变 $\hbar\omega$ 的整数倍，所以若格波的振动能量减少 $\hbar\omega$，则可以说它失去一个声子；反之，若振动能量增加 $\hbar\omega$，则获得一个声子。

同光子一样，声子遵守波色-爱因斯坦统计。因此，对于单位体积，频率 ω 至 $\omega+\mathrm{d}\omega$ 范围内的声子数是

$$N = g(\omega)n_q(\omega)\mathrm{d}\omega \tag{1.62}$$

函数 $g(\omega)$ 是声子态密度；$n_q(\omega)$ 是能态 $\hbar\omega$ 上的平均声子数，服从玻色-爱因斯坦分布：

$$n_q(\omega) = \frac{1}{\mathrm{e}^{\hbar\omega/k_BT}-1} \tag{1.63}$$

态密度 $g(\omega)$ 用类似于电子的方法(参见 1.1 节)来确定。在 q 空间中，一个状态占据的体积 V_q 和实际空间的有效体积 V 之间有下述关系：

$$V_q \cdot V = (2\pi)^3 \tag{1.64}$$

例如声学波，设 $\omega \approx Cq$，并设横向(二种模式)和纵向(一种模式)声子的速度皆为 C，则 q 和 $q+\mathrm{d}q$ 之间每单位体积的声学声子数是

$$g(q)\mathrm{d}q = 3 \cdot \left(\frac{4\pi q^2\mathrm{d}q}{8\pi^3}\right) \tag{1.65}$$

所以

$$g(\omega)\mathrm{d}\omega = g(q)\frac{\mathrm{d}q}{\mathrm{d}\omega}\mathrm{d}\omega = \frac{3\omega^2}{2\pi^2 C^3}\mathrm{d}\omega \tag{1.66}$$

声子的概念主要用于考虑格波对电子的散射。从量子力学的角度看，一个电子与声子碰撞时，是得到能量还是失去能量，归结于电子是吸收声子还是发射声子。

为简单明了，我们考虑电子和能量为 $\hbar\omega$ 的声子之间的相互作用。这样，当电子吸收一个声子时，能量的变化是 $\hbar\omega$，从低能态 E_1 向上跃迁到高能态 E_2，即 $E_2 - E_1 = \hbar\omega$，这一过程称为吸收(或受激吸收)。与这相反的过程是电子受到声子激发从 E_2 向下跃迁到 E_1，并且发射一个同样的声子，这一过程称为受激发射。上述两个过程互为逆过程。细致平衡原理表明，两个量子态之间的跃迁几率等于严格的逆跃迁的几率。因此，一个声子被能量 E_1 的电子所吸收的几率 b_{12} 应当等于同样的声子被能量为 E_2 的电子所发射的几率 b_{21}，即

$$b_{12} = b_{21} \tag{1.67}$$

相应地，吸收声子的总几率是 $b_{12}n_q(\omega)n_1$，发射声子的总几率是 $b_{21}n_q(\omega)n_2$，其中 $n_q(\omega)$ 是能量为 $\hbar\omega$ 的声子数，n_1 和 n_2 是能量分别为 E_1 和 E_2 的电子浓度。

然而，上面的论述没有考虑所谓自发发射过程。这是不存在声子激发时，电子从 E_2 自发向下跃迁到 E_1，并且发射一个能量为 $E_2 - E_1$(等于 $\hbar\omega$)的声子的过程。此过程可理解如下：由于半导体中总是有电子到处移动，即便原来晶格绝对静止，也会因为电子和晶格原子的电场相互作用，激励原子进入振动状态，即激发声子。这样发射一个声子自然与声子密度无关。自发发射的总几率为 $a_{21}n_2$，a_{21} 是每个电子发射声子的几率。

在平衡情况下，必须是吸收多少声子就发射多少声子，所以

$$b_{12}n_q(\omega)n_1 = b_{21}n_q(\omega)n_2 + a_{21}n_2 \tag{1.68}$$

利用式(1.67)得到

$$b_{12}\left[\frac{n_1}{n_2} - 1\right]n_q(\omega) = a_{21} \tag{1.69}$$

设电子按能量服从玻耳兹曼分布，则有

$$\frac{n_1}{n_2} = \mathrm{e}^{-(E_1-E_2)/k_\mathrm{B}T} = \mathrm{e}^{\hbar\omega/k_\mathrm{B}T} \tag{1.70}$$

同时，由式(1.63)得知 $n_q(\omega) = [\mathrm{e}^{\hbar\omega/k_\mathrm{B}T} - 1]^{-1}$，所以

$$b_{12} = a_{21} = b_{21} \tag{1.71}$$

由方程(1.68)和式(1.71)不难得出结论：

$$\begin{aligned} &\text{吸收声子几率} \propto n_q(\omega) \\ &\text{发射声子几率} \propto [n_q(\omega) + 1] \end{aligned} \tag{1.72}$$

最后，我们利用上述结论来确定电子和声子之间每次碰撞的能量损失。

设电子温度为 T_e，晶格温度为 T_0。根据上述，对于这样一个系统，吸收 $\hbar\omega$ 声子的几率是 $b_{12}n_q(\omega, T_0)n_1(T_\mathrm{e})$，发射(包括自发和受激发射)声子的几率是 $b_{21}[n_q(\omega, T_0)+1]n_2(T_\mathrm{e})$，且 $b_{12} = b_{21}$。能量的交换总是 $\hbar\omega$，所以平均说来每次碰撞的能量损失是

$$\begin{aligned} \Delta E &= \hbar\omega\frac{n_2(T_\mathrm{e})[n_q(\omega, T_0) + 1] - n_1(T_\mathrm{e})n_q(\omega, T_0)}{n_2(T_\mathrm{e})[n_q(\omega, T_0) + 1] + n_1(T_\mathrm{e})n_q(\omega, T_0)} \\ &= \hbar\omega\frac{\mathrm{e}^{\hbar\omega/k_\mathrm{B}T_0} - \mathrm{e}^{\hbar\omega/k_\mathrm{B}T_\mathrm{e}}}{\mathrm{e}^{\hbar\omega/k_\mathrm{B}T_0} + \mathrm{e}^{\hbar\omega/k_\mathrm{B}T_\mathrm{e}}} \end{aligned} \tag{1.73}$$

1.2.3 晶格散射

如果所有原子都在平衡位置，那么就有一个周期性电势，允许电子以有效质量自由运动。当任一原子位移时就产生一个电势的扰动，作用在电子上就会改变电子的动量和能量；另一方面，电子对移位的原子也有反作用，所以任何动量或能量的交换都是相互的。在良好的半导体中，单个位移原子所引起的散射是很弱的，只有全部位移原子的联合作用（即移过晶体的振动波，或者说格波）才是显著的。格波使晶体在许多晶格间距上发生形变，从图 1.1 表示的能带随晶格间距变化的图像不难看出，晶体的形变将引起电子势能的变化（形变势），它反射或散射电子波（形变势散射）。在非极性半导体（例如 Ge, Si）中，晶格散射为形变势散射，声学波引起的声学形变势散射的作用通常被认为大于光学形变势散射。对于谷内散射，电子和声子碰撞之后仍然在它原来的导带能谷内，因此，其动量的变化只是热动量 mv_{th} 的数量级（v_{th} 是对应电子温度 T_{e} 的平均热速度，为 $10^7\mathrm{cm/s}$ 量级），即碰撞中涉及的声子动量只可能是 mv_{th} 的量级，主要是长波。对于长声学波，由式(1.58)可知，声子的能量 $\hbar\omega \approx \hbar Cq$，故有

$$\frac{\hbar\omega}{k_{\mathrm{B}}T_{\mathrm{e}}} \sim \frac{Cmv_{\mathrm{th}}}{mv_{\mathrm{th}}^2} \sim \frac{C}{v_{\mathrm{th}}} \sim 10^{-2}(\text{数量级}) \tag{1.74}$$

因此电子在与声学声子碰撞时，尽管电子动量有很大变化，但电子得到或失去的能量部分是很小的，即碰撞是近似弹性的。在经典近似中，这种碰撞对应轻粒子和重粒子的碰撞。可以在 $\hbar\omega/k_{\mathrm{B}}T_{\mathrm{e}} \ll 1$ 或者 T_{e} 接近 T_0 条件下，通过式(1.73)计算声学声子每次碰撞的能量损失，结果为

$$\Delta E = \frac{(\hbar\omega)^2}{k_{\mathrm{B}}T_{\mathrm{e}}} \frac{T_{\mathrm{e}} - T_0}{T_{\mathrm{e}}} \tag{1.75}$$

对于谷间散射，电子从 k 空间的一个能谷变化到另一个能谷，相应的波矢量相当于两个能谷在 k 空间的间距。这样的声子接近于声学分支的顶部或光学分支的底部，它的能量总是和 $\hbar\omega_0$ 同数量级的（参阅图 1.9），这是任何一个声子可能取得的最高能量。若 T_{e} 接近 T_0，平均每次碰撞的能量损失仍由式(1.75)表示，只是 ω 应由 ω_0 代替。而当电子获得的能量使 T_{e} 远高于 T_0 时，每次碰撞的能量损失趋于一个常数值，$\hbar\omega_0 \tanh(\hbar\omega_0/k_{\mathrm{B}}T_{\mathrm{e}}) = \hbar\omega_{光}$，几乎是声子可能取得的最大能量。所以，谷间散射可能引起可观的能量损失。这一论述对于任何能量是 $\hbar\omega_0$ 量级的声子的相互作用都是正确的。就光学声子散射而言，当电子能量增大时，与声子每次碰撞的能量损失趋近一个常数 $\hbar\omega_{光}$（近似等于 $\hbar\omega_0$）。因此在非极性半导体中，光学声子在电子波的散射以及引起能量对晶格的损失中可能起重要作用。

在极性半导体（例如 GaAs）中，一类原子（As 原子）倾向于附着到另一类原子（Ga 原子）上，所以两类原子有相反的电荷，结果产生一个极化电场，把两类原子束缚在一起。两类原子的相对位移使平衡周期场变化，结果这一变化就直接作用在电子（或空穴）上，这种散射作用来自光学波（因为移动极性半导体中的极化原子的是光学波），称为极性光学波散射。因此，在极性半导体中，引起电阻和能量损失的主要是光学声子。根据上面的论述可以认为能量损失率的增加应该趋向于一个极限。实际并非如此。这可能是与极性光学声子碰到的有效几率的减小调和了每次碰撞的能量损失的增大。粗略地讲，在经典理论中相互作用的减少是极化场的局域化所引起的，高能量电子扫过原子之间极化电场的时间很短，以致能量损失率随能量的增加而减小。

1.3 载流子输运现象

通常采用漂移-扩散方程为核心的简单理论来近似描述半导体中载流子(电子和空穴)的稳态输运现象。本节除考虑载流子在电场作用下的漂移现象和在载流子浓度梯度作用下的扩散现象以外,还将考虑载流子在非平衡情形下的复合过程和产生过程;并且导出半导体器件工作过程所遵循的基本方程,包括电流密度方程(漂移-扩散方程)和连续性方程。最后,指出稳态输运理论的局限性,并简要讨论瞬态输运现象。

1.3.1 载流子漂移

1. 动量和能量弛豫时间

考虑一块均匀的 n 型半导体。在热平衡状态下,电子在半导体中朝各个方向迅速运动,其平均热运动速度 v_{th} 可由下式得到:

$$\frac{1}{2}m_n v_{th} = \frac{3}{2}k_B T \tag{1.76}$$

式中 m_n 为电子质量,T 为绝对温度。对室温(300K)的硅和砷化镓,v_{th} 约为 10^7cm/s。可以把单个电子的热运动看成是电子与晶格原子、杂质原子及其他散射中心的一系列随机碰撞,在足够长的时间间隔内电子随机运动的净位移为零。两次碰撞之间的平均时间称为平均自由时间 τ,不难验证 τ 与粒子在上一次碰撞之后的时间 t 至 $t+\mathrm{d}t$ 之间的散射几率 $P(t)\mathrm{d}t$ 有下述关系

$$P(t)\mathrm{d}t = \frac{1}{\tau}\mathrm{e}^{-t/\tau}\mathrm{d}t \tag{1.77}$$

相继两次碰撞之间的平均距离称为平均自由程 λ,类比于平均自由时间,我们说上一次碰撞后的 x 至 $x+\mathrm{d}x$ 距离上的碰撞几率为

$$P(x)\mathrm{d}x = (1/\lambda)\mathrm{e}^{-x/\lambda}\mathrm{d}x \tag{1.78}$$

应该注意到,如果粒子的平均自由程与速度 v 无关,则平均自由时间与速度有关($\tau=\lambda/v$)。普遍地讲,τ 与粒子的能量 E 有关,即 $\tau=\tau(E)$。

当有一个小的电场$\mathscr{E}$加到半导体样品上时,每个电子在电场力 $-q\mathscr{E}$ 的作用下,在相继两次碰撞之间会被电场加速(与电场方向相反)。因此在电子热运动上会叠加一个附加的速度分量,这个附加的速度分量称为漂移速度。

如果电子自由运动时从电场中获的动量在碰撞时全部失去,则在两次碰撞之间电子作自由运动时,电场给电子的冲量(力×时间)等于该期间内电子获得的动量,即可得到稳态情形下电子的漂移速度。电场给电子的冲量为 $-q\mathscr{E}\tau$,电子获得的动量为 $m_n v_n$,因此有

$$-q\mathscr{E}\tau = m_n v_n \tag{1.79}$$

式中,τ 为平均自由时间,v_n 为电子的漂移速度。

实际上碰撞并不像上面假定的那样理想,一个改进的模型是,载流子(电子或空穴)在碰撞之后一定数量的动量和能量保留下来,而由于相碰的结果又加上一定的无规量。为简单起见,我们假定电子每次碰撞后动量的不变分数是 b,并且与原来动量同方向;而附加的无规动量是 $\boldsymbol{w}$。原来动量的其余部分传给晶格或者失去。这样,取 $\boldsymbol{P}_n$ 为在第 n 次相碰前一个电子的动

量，在第 n 次相碰后就变为 $b\boldsymbol{P}_n + \boldsymbol{w}$，电场力 $q\mathscr{E}$ 正好沿着电场方向增加动量。所以，在第 $n+1$ 次相碰之前，在第 n 次相碰后的一段时间 t 内，我们有

$$P_{n+1} = b\boldsymbol{P}_n + \boldsymbol{w} + q\mathscr{E}t \tag{1.80}$$

我们注意到电子的平均动量 $\langle \boldsymbol{P}_n \rangle$ 与已经进行的碰撞次数 n 无关，所以 $\langle \boldsymbol{P}_n \rangle$ 恰好就是平均动量

$$\langle \boldsymbol{P}_n \rangle = m\boldsymbol{v} \tag{1.81}$$

动量弛豫时间

设粒子的平均自由时间为 τ，利用(1.77)式引入的自由时间分布来平均(1.80)式的两边，我们得到

$$m\boldsymbol{v} = bm\boldsymbol{v} + q\mathscr{E}\tau$$

$$m\boldsymbol{v} = q\mathscr{E}\tau_m, \quad \tau_m = \frac{\tau}{1-b} \tag{1.82}$$

由于 $\boldsymbol{w}$ 在方向上是无规则分布的，所以已令 $\langle \boldsymbol{w} \rangle = 0$，方程(1.82)可作如下解释。如果每次碰撞失去的动量分数是 $(1-b)$，则失去全部动量的有效碰撞次数为 $1/(1-b)$。参数 τ_m 就可以看作是丧失动量碰撞之间的有效平均时间。对于各向同性散射，散射过程是对称的，参量 b 的平均值是零，这时 τ_m 和 τ 这两个时间常数没有区别。

我们知道，一个力的作用能产生一个动量变化，如果把碰撞作用看作一个和电场方向相反、强度为 $-\boldsymbol{P}/\tau_m$ $(\boldsymbol{P} = m\boldsymbol{v})$ 的阻力，则带电粒子受到电场力 $q\mathscr{E}$ 和这个阻力的共同作用，由稳态情形下动量变化率 $\mathrm{d}\boldsymbol{P}/\mathrm{d}t = 0$ 的条件，就得到方程(1.82)。而在没有电场时，可以预料

$$\frac{\mathrm{d}\boldsymbol{P}}{\mathrm{d}t} = -\boldsymbol{P}/\tau_m \tag{1.83}$$

因此当去掉电场时，粒子动量平均以时间常数率 τ_m 衰减或弛豫到零。为此，把 τ_m 称作动量弛豫时间。

能量弛豫时间

我们应用同样的模型来讨论电场 $\mathscr{E}$ 到粒子的能量转移。我们从方程(1.80)的两边取平方出发

$$\boldsymbol{P}_{n+1}^2 = b^2\boldsymbol{P}_n^2 + \boldsymbol{w}^2 + (q\mathscr{E}t)^2 + 2b\boldsymbol{P}_n \cdot q\mathscr{E}t + 2b\boldsymbol{P}_n \cdot \boldsymbol{w} + 2\boldsymbol{w} \cdot q\mathscr{E}t \tag{1.84}$$

可以使用像对(1.80)式同样的平均过程，但必须注意，虽然包含单独 $\boldsymbol{w}$ 的项的平均值是零，然而 $\boldsymbol{w}^2$ 项的平均值不是零。现在，$\langle \boldsymbol{P}_n^2/2m \rangle = E$ 是每个粒子的平均能量，既然是平均值，那就与 n 的值无关。因此，利用这个式子以及(1.82)式，在进行整理后得到

$$(1-b^2)E = \frac{\langle \boldsymbol{w}^2 \rangle}{2m} + \frac{(q\mathscr{E})^2}{m}\tau_m\tau \tag{1.85}$$

在平均过程中还用到下式：

$$\langle t^2 \rangle = \int_0^\infty \frac{t^2}{\tau}\mathrm{e}^{t/\tau}\mathrm{d}t = 2\tau^2 \tag{1.86}$$

如果电场是零，则稳态情形下粒子与晶格处于某一温度 T_0 的热平衡状态，此时粒子的能量应该相当于 $\frac{3}{2}k_B T_0 = E_0$，而必然有

$$\frac{\langle \boldsymbol{w}^2 \rangle}{2m} = (1-b^2)E_0 \tag{1.87}$$

方程(1.85)则变成

$$E - E_0 = q\mathscr{E} \cdot v\tau_E, \text{ 其中 } \tau_E = \frac{\tau}{1 - b^2} \tag{1.88}$$

在更普遍的情况下,参量 b 可能是依赖于方向的变数,此时在这里使用的值应该是平方值的平均。

正如动量弛豫时间一样,如果去掉电场$\mathscr{E}$,则粒子能量以时间常数 τ_E 弛豫到热平衡值 E_0。τ_E 称作能量弛豫时间。

根据上面的讨论,稳态情形下有

$$\tau_m(E) = \frac{mv}{q\mathscr{E}}, \quad \tau_E(E) = \frac{E - E_0}{q\mathscr{E}v} \tag{1.89}$$

为计算 $\tau_m(E)$和 $\tau_E(E)$,可以用数值方法模拟半导体中的一个电子(或空穴)在电场和声子、杂质等散射机制共同作用下,在足够长时间内的运动,得到有关物理量的统计平均值,例如,漂移速度 v 和平均能量 E 等(蒙特-卡罗方法)。这些宏观物理量显然和电场强度$\mathscr{E}$有关。图1.12和图 1.13 分别为利用蒙特-卡罗方法和方程(1.89)计算得到的 Si 和 GaAs 的动量弛豫时间和能量弛豫时间。由图可见,能量弛豫时间比动量弛豫时间大一个数量级。只要考虑到 $\tau_m(E)$取决于各种散射(包括弹性的和非弹性的),而 $\tau_E(E)$只取决于非弹性散射,就不难理解这一结果。

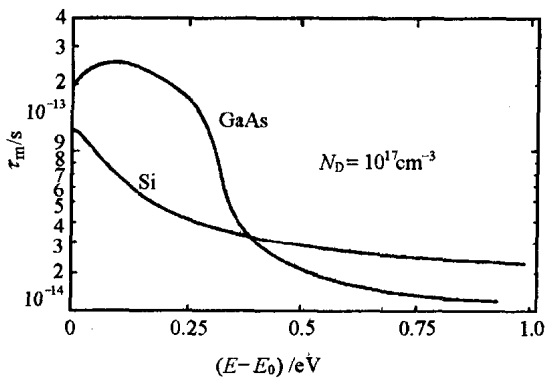


图 1.12 Si 和 GaAs 的动量弛豫时间(温度为 300K)

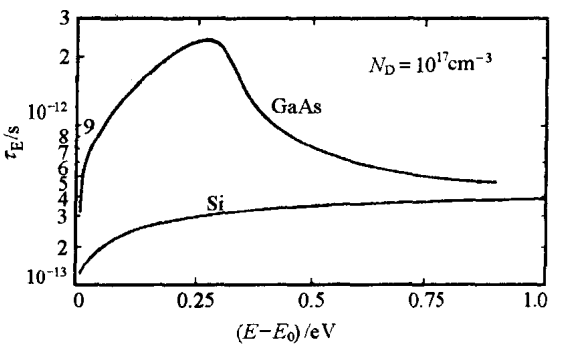


图 1.13 Si 和 GaAs 的能量弛豫时间(温度为 300K)

2. *迁移率*

通过动量平衡方程 $q\mathscr{E}\tau_m = mv$[参见(1.89)式],直接给出迁移率 μ 的概念:

$$\mu = \frac{q\tau_m}{m} \tag{1.90}$$

其中 m 是载流子有效质量,τ_m 是动量弛豫时间。如前所述,如果散射是对称的,则 τ_m 和 τ 这两个时间没有区别。因为常常是这种情况,我们倾向在初步的处理中认为两个概念之间没有区别。

平均自由时间 τ(或动量弛豫时间 τ_m)是粒子平均能量 E 的函数,只是通过 E 间接依赖于电场 $\mathscr{E}$,$\mathscr{E}$特别小时与电场无关,这时漂移速度和电场成正比,比例常数就是低场迁移率。低电场时,对电子有

$$v_n = -\mu_n \mathscr{E} \tag{1.91}$$

负号表示电子的运动方向与电场方向相反；式中 $\mu_n = q\tau/m_n$ 是电子迁移率。对空穴有

$$v_p = \mu_p \mathscr{E}, \text{ 其中 } \mu_p = \frac{q\tau}{m_p} \tag{1.92}$$

μ_p 为空穴迁移率。

由上述有关的表示式可见，迁移率与碰撞的平均自由时间直接有关，而碰撞又取决于各种散射机构。两种最重要的散射机构是晶格散射和电离杂质散射。晶格散射起因于晶格原子在绝对零度以上的热振动，我们在 1.2 节作了比较详细的讨论。当电场 $\mathscr{E}$ 十分小时，载流子能量偏离热平衡值很小，主要受声学声子散射，即长声学波散射。计算表明，长声学波散射是各向同性的，对载流子的散射几率（平均碰撞频率，即 $1/\tau$）正比于格波振幅的平方，亦即正比于其振动能量 $k_B T \approx n_q \hbar\omega$（$n_q$ 为每一种振动模式的平均声子数，表示式参见(1.63)式）。此外，散射几率还应正比于终态的态密度（$\sim E^{1/2}$）。因此有

$$\tau_L \propto \frac{E^{-1/2}}{k_B T} \tag{1.93}$$

可见对于长声学波散射，散射几率除了依赖于晶格振动的强度外（依赖于温度），还依赖于载流子自身能量 E，温度愈高，载流子平均能量亦应愈高，而有 $E \propto T$，从而 $\tau_L \propto T^{-3/2}$，这导致晶格散射迁移率 μ_L 按 $T^{-3/2}$ 比例减小。

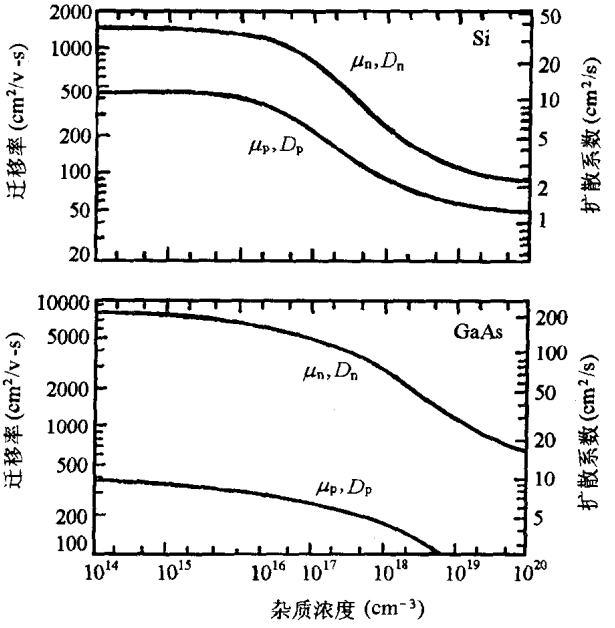


图 1.14　Si 和 GaAs 中迁移率、扩散系数和杂质浓度的关系（温度为 300K）

载流子经过电离的施主或受主杂质时，将受到库仑力的作用，运动方向发生偏折（卢瑟福散射），亦即发生杂质散射。杂质散射的几率反比于载流子能量的平方（$\sim E^{-2}$）和正比于终态的态密度（$\sim E^{1/2}$），因而有 $\tau_I \propto E^{3/2}$，即 $\tau_I \propto T^{3/2}$。这说明杂质散射在高温下显著减弱。因为

在高温时，载流子运动加快，停留在杂质原子附近的时间较短，因此受到的有效散射较小。另一方面，杂质散射的几率也正比于电离杂质的总浓度，即带负电离子和带正电离子浓度的总和。这样，杂质散射迁移率 μ_I 按 $T^{3/2}N_I^{-1}$ 变化（N_I 为杂质总浓度）。

在同时存在几种散射机构并且彼此"独立"（即一种散射机构的存在并不影响其他散射机构的作用方式）时，总的几率应为各散射几率之和。对于晶格散射和电离杂质散射，可以得到

$$\frac{1}{\tau} = \frac{1}{\tau_L} + \frac{1}{\tau_I} \tag{1.94}$$

或

$$\frac{1}{\mu} = \frac{1}{\mu_L} + \frac{1}{\mu_I} \tag{1.95}$$

图 1.14 表示室温下硅和砷化镓实测的迁移率与杂质浓度的关系。迁移率在低杂质浓度下达到一个最大值，这对应于晶格散射的极限。一定温度（例如 300K）下，因杂质散射作用的增强，迁移率随杂质浓度的增大而减小，最后在高杂质浓度下趋近一个最小值。可以看到电子迁移率比空穴迁移率要大，这主要是因为电子的有效质量较小的缘故。

3. 电阻率

现在讨论均匀半导体材料的电导。低电场时，可以利用迁移率把载流子在电场作用下输运产生的电流（漂移电流）表示出来。电子的漂移电流密度可表示为

$$J_n = -qnv_n = qn\mu_n\mathscr{E} \tag{1.96}$$

n 为电子浓度。与此类似，对于空穴有

$$J_p = qp\mu_p\mathscr{E} \tag{1.97}$$

在外加电场 $\mathscr{E}$ 作用下，流过半导体样品的总电流密度是电子电流密度分量和空穴电流密度分量之和：

$$J = J_n + J_p = (qn\mu_n + qp\mu_p)\mathscr{E} \tag{1.98}$$

括号中的量称为电导率，即

$$\sigma = (qn\mu_n + qp\mu_p) \tag{1.99}$$

电子和空穴对电导率的贡献是简单相加。

半导体的电阻率 ρ 是 σ 的倒数，即

$$\rho = \frac{1}{\sigma} = \frac{1}{q(n\mu_n + p\mu_p)} \tag{1.100}$$

对于掺杂的非本征半导体来说，在式(1.98)或式(1.99)中，通常只有一个电流分量是主要的，因为两种载流子的浓度相差好几个数量级。因此，对 n 型半导体，式(1.100)可简化为

$$\rho = \frac{1}{qn\mu_n} \tag{1.101}$$

对 p 型半导体，式(1.100)可简化为

$$\rho = \frac{1}{qp\mu_p} \tag{1.102}$$

图 1.15 表示硅和砷化镓(300K)的实测电阻率与杂质浓度的函数关系。在这个温度和低杂质浓度情况下，具有浅能级的施主或受主杂质将全部电离。此时，载流子浓度等于杂质浓度。若半导体电阻率已知，可以从这些曲线求得杂质浓度，反之，已知杂质浓度，也可求出电阻

率。

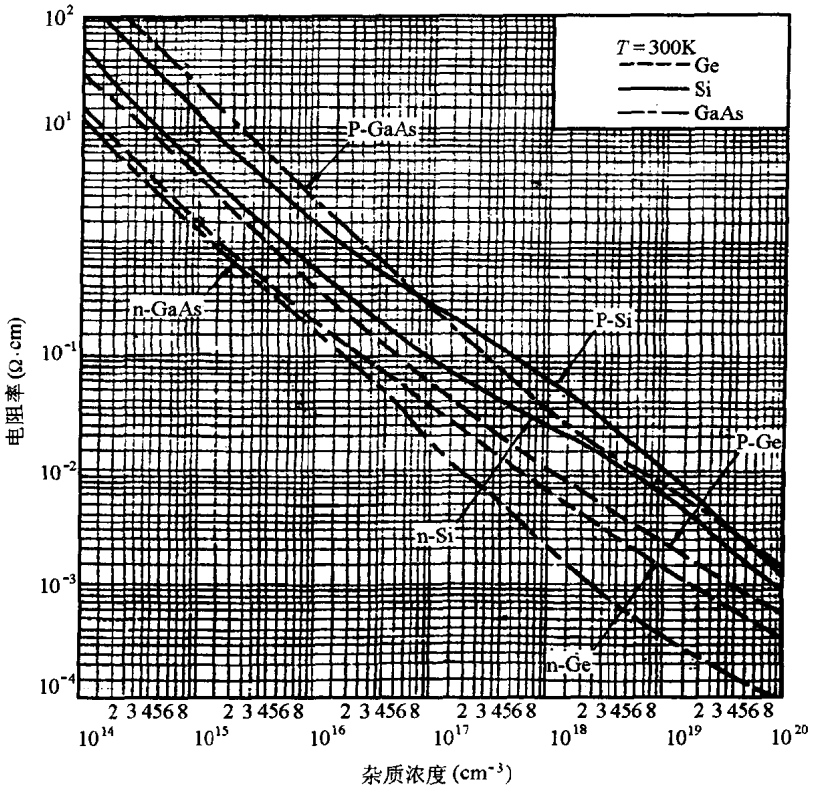


图 1.15　Si 和 GaAs 的电阻率和浓度的关系

此外 Si 和 Ge 中，导带有多个旋转椭球对称的等价能谷，电子电流应该包括所有这些能谷中电子的贡献。如果引入电导有效质量的概念，则材料的输运行为和各向同性的半导体类似，即电子漂移电流密度形式仍可表示为

$$J_n = qn\mu_n\mathscr{E} \tag{1.103}$$

但迁移率表示式中的有效质量为电导有效质量，即

$$\mu_n = \frac{q\tau}{m_\sigma} \tag{1.104}$$

式中 m_σ 为电导有效质量。Si 的导带在 $\boldsymbol{K}$ 空间的<100>方向共有 6 个等价能谷，m_σ 满足下式：

$$\frac{1}{m_\sigma} = \frac{1}{3}\left[\frac{1}{m_l} + \frac{2}{m_t}\right] \tag{1.105}$$

m_l，m_t 分别为纵向、横向有效质量。

1.3.2　载流子扩散

上一小节讨论了载流子在电场作用下输运产生的电流，即漂移电流。若半导体材料中载流子浓度存在空间上的变化，载流子倾向于从高浓度区流向低浓度区，这种由于扩散作用引起的电流称为扩散电流。

扩散是通过载流子的热运动实现的。由于热运动,不同区域不断进行着载流子的交换,若载流子不均匀分布,这种交换就会引起载流子的宏观流动。因此,扩散流的大小与载流子的绝对数量没有直接联系。为了进一步理解扩散过程,假设电子浓度沿 x 方向变化,半导体的温度是均匀的,因此电子的平均热能不随 x 变化,只是浓度 $n(x)$ 随 x 变化。我们考虑单位时间内通过 $x=0$ 平面单位面积上的电子数。在一定温度下,电子以热运动速度 v_{th},平均自由程 λ ($\lambda=v_{th}\tau$, τ 为平均自由时间)作无规则热运动。在 $x=-\lambda$(即左边距离 $x=0$ 为一个平均自由程处)的电子向左或向右的机会相等;在一个平均自由时间 τ 内,有一半电子穿过 $x=0$ 的平面。因此,单位时间内从左向右通过 $x=0$ 平面单位面积的电子平均流量 F_1 为

$$F_1=\frac{\frac{1}{2}n(-\lambda)}{\tau}=\frac{1}{2}n(-\lambda)v_{th} \tag{1.106}$$

同样,单位时间内在 $x=\lambda$ 处的电子从右向左通过 $x=0$ 平面单位面积的平均流量 F_2 为

$$F_2=\frac{1}{2}n(\lambda)v_{th} \tag{1.107}$$

这样,单位时间内电子从左向右的净流量

$$F=F_1-F_2=\frac{1}{2}v_{th}[n(-\lambda)-n(\lambda)] \tag{1.108}$$

将随空间变化的电子浓度在 $x=\pm\lambda$ 处展开成泰勒级数,取头两项作为近似,可得

$$F=-v_{th}\lambda\frac{dn}{dx}\equiv-D_n\frac{dn}{dx} \tag{1.109}$$

式中 $D_n=v_{th}\lambda$ 称为扩散系数。每个电子所带电荷为 $-q$,因此电子流动产生的电流为

$$J_n=-qF=qD_n\frac{dn}{dx} \tag{1.110}$$

扩散电流正比于电子浓度对空间的导数,它是在有浓度梯度时由载流子的无规热运动引起的。随着 x 的增加电子浓度梯度是正时,电子将向负 x 方向扩散,此时电流是正的,即电流方向与电子扩散的方向相反。

1.3.3 电流密度方程(漂移-扩散方程)

通常忽略热电子效应,即假定电子温度和晶格温度相同(热平衡近似)。在这种情形下,当电场和浓度梯度同时存在时,半导体内任一点的电流是漂移分量和扩散分量之和:

$$J_n=q\mu_n n\mathscr{E}+qD_n\frac{dn}{dx} \tag{1.111}$$

扩散系数和迁移率二者受一共同的因素 τ 制约,因而有一定的必然联系。现在,我们寻求热平衡时它们之间的关系。热平衡时,各处费米能级在同一水平上,载流子浓度的高低反映在带边与费米能级距离的大小上。带边的弯曲意味着存在电场,它的作用在于产生适当的漂移电流与浓度梯度引起的扩散电流相抵消。我们注意到,存在电场 $\mathscr{E}$ 时,每个电子受电场力 $-q\mathscr{E}$ 作用,电场力等于势能梯度的负值,即 $-q\mathscr{E}=-$(电子势能梯度)。本章开头我们就曾经指出,导带底 E_c 相当于电子的势能,因而有

$$\mathscr{E}=\frac{1}{q}\frac{dE_c}{dx} \tag{1.112}$$

我们还注意到,热平衡电子浓度(参见(1.27)式)为

$$n = N_c F_{1/2}(\eta_c) \tag{1.113}$$

其中 $\eta_c=(E_F-E_c)/k_BT$，所以电子浓度是能量差 E_F-E_c 的函数，即 $n=n(E_F-E_c)$。注意到这两个方面以及热平衡时电流为零，对于方程(1.111)令 $J_n=0$，可以证明

$$\mu_n = qD_n \frac{d(\ln n)}{dE_F} \tag{1.114}$$

式(1.114)表示广义的爱因斯坦关系，不论半导体简并或非简并它都适用。非简并时，$n=N_c e^{\eta_c}$，我们得到熟知的爱因斯坦关系：

$$D_n = \left[\frac{k_BT}{q}\right]\mu_n \tag{1.115}$$

有电流时，电子分布总在一定程度上偏离平衡。只要相对平衡分布的偏离不是很大，我们可以定义与位置有关的费米能级(准费米能级)$E_F(x)$，把方程(1.113)推广到非平衡情形，并且认为爱因斯坦关系式仍然成立。用准费米能级 $E_F(x)$和广义爱因斯坦关系[式(1.114)]，可以把电流方程(1.111)改写为

$$\boldsymbol{J}_n = \mu_n n \frac{dE_F(x)}{dx} \tag{1.116}$$

在三维情形下，梯度符号$\frac{d}{dx}$应该用$\nabla=\frac{\partial}{\partial x}\mathrm{i}+\frac{\partial}{\partial y}\mathrm{j}+\frac{\partial}{\partial z}\mathrm{k}$代替，i、j、k 分别为沿 x、y、z 轴方向的单位矢量。这样，电子电流密度为

$$\boldsymbol{J}_n = q\mu_n n\mathscr{E} + qD_n\nabla n \tag{1.117}$$

或

$$\boldsymbol{J}_n = \mu_n n \nabla E_F(r) \tag{1.118}$$

对空穴电流可得与电子电流类似的表达式，在三维情形下为

$$\boldsymbol{J}_p = q\mu_p p\mathscr{E} - qD_p\nabla p \tag{1.119}$$

或

$$\boldsymbol{J}_p = \mu_p p \nabla E_F(r) \tag{1.120}$$

式(1.119)中的扩散项取负号，是由于空穴浓度梯度(一维情形是 dp/dx)为正时，空穴将向负 x 方向扩散，由此引起的空穴电流也流向负 x 方向。扩散系数 D_p 和迁移率 μ_p 之间同样也存在爱因斯坦关系

$$\mu_p = -qD_p \frac{d(\ln p)}{dE_F} \tag{1.121}$$

非简并情形下，式(1.126)简化为下述熟知的表示式：

$$D_p = \left[\frac{k_BT}{q}\right]\mu_p \tag{1.122}$$

总的传导电流密度是电子电流密度与空穴电流密度之和：

$$\boldsymbol{J} = \boldsymbol{J}_n + \boldsymbol{J}_p \tag{1.123}$$

式(1.123)以及关于 $\boldsymbol{J}_n$，$\boldsymbol{J}_p$ 的表达式构成电流密度方程。这些方程对分析器件在低电场下的工作是非常重要的。但强电场下迁移率与电场有关，$\mu_n\mathscr{E}$ 及 $\mu_p\mathscr{E}$ 应该用漂移速度 $\boldsymbol{v}_n$ 及 $\boldsymbol{v}_p$ 代替(电场足够高时，v_n，v_p 趋近于饱和值 v_s)。另一方面，电子(或空穴)的浓度梯度很大时，扩散流偏离式(1.109)表示的扩散定律(Fick 定律)；否则，扩散流中电子的速度 $n^{-1}D_n\nabla n$ 可

以无限大,这显然是不符合实际的。

1.3.4 过剩载流子及其产生与复合

在热平衡下,关系式 $pn = n_i^2$ 成立。若有过剩载流子被引入半导体,使得 $pn > n_i^2$,这时半导体处于非平衡状态。引入过剩载流子的过程称为载流子注入。可以用多种方法注入载流子,其中包括光激发和对 p-n 结加正向偏置(后者将在第二章中讨论)。在光激发的情形下,用光照射半导体,若光子能量 $h\nu$ 大于半导体禁带宽度 E_g(这里 h 为普朗克常量,ν 为光的频率),则光子被半导体吸收并产生一个电子-空穴对。光激发增加了电子和空穴的浓度,使其超过平衡值,这些新增加的载流子叫过剩载流子。过剩载流子也常称为非平衡载流子。这里"非平衡"仅指数量上偏离平衡值,其能量分布仍可与平衡分布相同。相反,在有些情形下,载流子的浓度虽然未发生显著变化,但其能量分布却可以是非平衡的,强场下的热电子就是这种情况。

1. 准费米能级

我们知道,热平衡时电子和空穴的分布由费米函数给出;一旦知道费米能级,分布函数就被确定。显然,如果过剩电子和空穴注入半导体,就不能用同样的函数来描述状态的占据几率。在一定的假设下,电子和空穴的占据几率用准费米能级描述。这些假设是:(1) 电子实际上在导带内热平衡,空穴在价带内热平衡。这意味着载流子既不能从晶格原子得到能量,也不会失去能量。(2) 电子-空穴复合时间比起电子在导带内和空穴在价带内达到平衡的时间来说要长得多。

在许多实际问题中,在同一能带内达到平衡的时间大约小于 10^{-10}s,而电子-空穴复合时间通常是 μs 数量级。这样,上述假设通常可以满足。在这种情形下,准平衡的电子和空穴可以由电子费米函数 $f_n(E)$和空穴费米函数 $f_p(E)$表示,并有

$$n = \int_{E_c}^{\infty} f_n(E) g_c(E) \mathrm{d}E \tag{1.124}$$

$$p = \int_{-\infty}^{E_v} f_p(E) g_v(E) \mathrm{d}E \tag{1.125}$$

其中

$$f_n(E) = \frac{1}{e^{(E-E_{Fn})/k_B T} + 1} \tag{1.126}$$

E_{Fn}为电子准费米能级。若 $f_v(E)$是价带中的电子占据率,空穴占据率就是

$$f_p(E) = 1 - f_v(E) = 1 - \frac{1}{e^{(E-E_{FP})/k_B T} + 1}$$

$$= \frac{1}{e^{(E_{FP}-E)/k_B T} + 1} \tag{1.127}$$

E_{FP}为空穴准费米能级。

2. 直接复合

每当热平衡状态受到破坏($pn \neq n_i^2$),就会存在一个使系统恢复平衡($pn = n_i^2$)的过程。在注入过剩载流子的情况下,恢复平衡的机制是导带中的过剩电子逐渐回到价带之中。这个过程称为复合。

电子和空穴的复合总是通过具体过程实现的。就复合过程中电子、空穴所经历的状态来说，复合过程可分为直接复合和间接复合。在直接复合过程中，电子直接跃迁至价带中的某一空状态。而在间接复合过程中，电子在跃迁到价带某一空状态之前还要经历某一(或某些)中间状态，例如电子先跃迁至某一杂质(缺陷)能级，然后再向价带跃迁。在复合之前，电子和空穴也可以由于库仑相互作用，先结合在一起，形成一种束缚态——激子，再通过激子的淬灭实现复合(激子复合)。在复合的每一跃迁过程中，不管是电子直接跃迁至价带、跃迁至杂质能级，还是激子的形成和淬灭，都能释放能量和动量。释放出来的能量可作为光子发射出去，也可以热能(发射声子，且通常是多声子过程，因为通常跃迁中释放的能量比单个声子的能量大得多)的形式传给晶格，这取决于复合过程的性质。发射光子的复合过程称为辐射复合，否则称为非辐射复合。

直接禁带中半导体的导带底和价带顶在 $\boldsymbol{K}$ 空间的同一位置，电子和空穴通过禁带复合不需涉及晶体动量变化(即不需要吸收或发射声子)。所以，在直接半导体中引入过剩载流子时，电子和空穴直接复合的几率很高。可以预料，单位时间和单位体积内复合的电子-空穴对数正比于电子浓度 n 和空穴浓度 p：

$$R = \gamma np \tag{1.128}$$

R 为复合速率；γ 称为直接复合系数，它实际上是对于各种能量的电子和空穴的平均值。在非简并情形下，电子和空穴有相同的能量分布，γ 应与电子和空穴的浓度无关。

热平衡时，产生率(单位时间和单位体积内产生的电子-空穴对数) G_{th} 和复合率 R_{th} 必定相等，以使载流子浓度保持常数，并且维持条件 $pn = n_{\mathrm{i}}^2$ 成立。

因此，对 n 型半导体有

$$G_{\mathrm{th}} = R_{\mathrm{th}} = \gamma n_{\mathrm{n0}} p_{\mathrm{n0}} \tag{1.129}$$

式中 n_{n0}，p_{n0} 分别为 n 型半导体热平衡时的电子浓度和空穴浓度。

有外界激发(例如光照)时，半导体的载流子浓度将超过其平衡值。此时，复合率和产生率分别为

$$R = \gamma n_{\mathrm{n}} p_{\mathrm{n}} = \gamma(n_{\mathrm{n0}} + \Delta n)(p_{\mathrm{n0}} + \Delta p) \tag{1.130}$$

和

$$G = G_{\mathrm{L}} + G_{\mathrm{th}} \tag{1.131}$$

G_{L} 表示光激发产生率。式中 Δn 和 Δp 为过剩载流子浓度，并有 $\Delta n = \Delta p$ 以保持整体的电中性。

空穴浓度的净变化率为

$$\frac{\mathrm{d}p_{\mathrm{n}}}{\mathrm{d}t} = G - R = G_{\mathrm{L}} + G_{\mathrm{th}} - R \tag{1.132}$$

在稳态下，$\frac{\mathrm{d}p_{\mathrm{n}}}{\mathrm{d}t} = 0$，由上式可得

$$G_{\mathrm{L}} = R - G_{\mathrm{th}} \equiv U \tag{1.133}$$

式中 U 为净复合率。把式(1.129)，(1.130)代入式(1.133)得到

$$U \approx \gamma(n_{\mathrm{n0}} + p_{\mathrm{n0}} + \Delta p)\Delta p \tag{1.134}$$

对于小注入，$\Delta p \ll n_{\mathrm{n0}}$，且对 n 型半导体有 $p_{\mathrm{n0}} \ll n_{\mathrm{n0}}$，故式(1.134)可简化为

$$U \approx \gamma n_{n0}\Delta p = \frac{p_n - p_{n0}}{1/(\gamma n_{n0})} \tag{1.135}$$

因此,净复合率正比于过剩少数载流子浓度,比例常数$\frac{1}{\gamma n_{n0}}$称为过剩少数载流子的寿命 τ_p。式(1.135)可写为

$$U = \frac{p_n - p_{n0}}{\tau_p} \tag{1.136}$$

其中

$$\tau_p = \frac{1}{\gamma n_{n0}} \tag{1.137}$$

寿命的物理意义可用外界激发因素(例如光照)撤除以后过剩载流子消失的情形来说明。设照射 n 型半导体的光源在 $t=0$ 时刻关断,则由式(1.132)可知,关断光照($G_L=0$)以后,空穴浓度的净变化率为

$$\frac{dp_n}{dt} = G_{th} - R = -U = -\frac{p_n - p_{n0}}{\tau_p} \tag{1.138}$$

求解方程(1.138),得到过剩载流子浓度按指数规律衰减:

$$\Delta p_n(t) = (\Delta p_n)_0 e^{-t/\tau_p} \tag{1.139}$$

$(\Delta p_n)_0$ 为 $t=0$ 时的过剩载流子浓度。由上式容易证明 τ_p 就等于过剩载流子的平均存在时间,即寿命。在 t 和 $t+dt$ 之间复合掉的过剩载流子为

$$d(\Delta p_n) = -\frac{(\Delta p_n)_0}{\tau_p} e^{-t/\tau_p} dt \tag{1.140}$$

因此,过剩载流子的平均存在时间为

$$\langle t \rangle = \frac{\int_0^\infty (\Delta p_n)_0 \frac{t}{\tau_p} e^{-t/\tau_p} dt}{(\Delta p_n)_0} = \tau_p \tag{1.141}$$

可见τ_p 就是过剩载流子的寿命。

τ_p 的倒数代表过剩载流子的复合几率 P:

$$P = \frac{1}{\tau_p} \tag{1.142}$$

它可理解为每个过剩载流子在单位时间内发生复合的平均次数。

3. 间接复合

像硅这类间接禁带半导体,导带底和价带顶不在 k 空间的同一位置,即导带底的电子相对价带顶的空穴有非零的晶体动量变化。因此,不与晶格相互作用而动量和能量都守恒的直接跃迁是不可能的。这类半导体中的主要复合过程是通过禁带中局域能态(杂质或缺陷能级)的间接跃迁。这些局域能态相当于导带与价带之间的"跳板",跃迁几率依赖于导带以及价带边缘与"跳板"之间的能量差,这些中间能级的存在会显著加快复合过程。这些中间能态也称复合中心,或产生-复合中心,因为这些能态对自由载流子既能起复合作用,又能起产生作用。

几乎所有半导体物理教科书都对间接复合理论做了详细叙述,我们这里予以省略。根据这一理论,通过单能级复合中心的净复合率为

$$U = \frac{pn - n_i^2}{\tau_P(n + n_t) + \tau_n(p + p_t)} \tag{1.143}$$

其中，

$$n_t = n_i e^{(E_t - E_i)/k_B T} \tag{1.144}$$

$$p_t = n_i e^{(E_i - E_t)/k_B T} \tag{1.145}$$

E_t 是复合中心的能级位置。

为了便于理解式(1.143)的主要特征，我们先考虑一个极限情况，即小注入下的 n 型半导体，此时 $n_n \gg p_n$。并假定复合中心靠近禁带中线，使 $n_n \gg n_t$。因此 U 可近似为

$$U \approx \frac{p_n - p_{no}}{\tau_p} \tag{1.146}$$

这个表达式与式(1.136)的形式相同，所以 τ_p 为相应的 n 型半导体中空穴的寿命(平均存在时间)。

平均存在时间可通过俘获截面 σ 来描述。我们设想空穴具有截面积 σ_p，则它在单位时间内扫过的体积是 $\sigma_p v_{th}$(v_{th}是热运动速度)。如果复合中心在该体积内，空穴将被它俘获。复合中心可看作直径是零，浓度为 N_t。则单位时间内每个空穴被俘获 $N_t \sigma_p v_{th}$ 次，即平均存在时间为

$$\tau_p = \frac{1}{N_t \sigma_p v_{th}} \tag{1.147}$$

σ_p 是复合中心对空穴的有效俘获截面。俘获截面的概念，可以完全等效地以另一种方式来实现，这便是假设复合中心具有有效面积，而粒子(空穴)的直径是零。每单位时间内的平均俘获次数仍然是 $N_t \sigma_p v_{th}$。因此俘获截面在粒子和复合中心之间具有互易的性质，但是不能同时指定两者都有有效面积，并且不能设想俘获截面具有粒子或复合中心的实际横截面积的物理意义。

对于 p 型半导体中电子的寿命，可以得到类似的表达式

$$\tau_n = \frac{1}{N_t \sigma_n v_{th}} \tag{1.148}$$

σ_n 是复合中心对电子的有效俘获截面。

室温下，n 型硅中寿命 τ_p 的典型值约为 0.3μs，P 型硅中寿命 τ_n 为 1.0μs。寿命与多数载流子浓度无关。这是因为 n 型半导体中存在大量的电子，只要有一个空穴被复合中心俘获，立刻就有一个电子被该中心俘获，从而完成复合过程。因此，复合过程中限制速率的是少数载流子的俘获。

将式(1.144)，(1.145)，(1.147)和(1.148)代入式(1.143)，有

$$U = \frac{v_{th}\sigma_n\sigma_p N_t (pn - n_i^2)}{\sigma_p [p + n_i e^{(E_i - E_t)/k_B T}] + \sigma_n [n + n_i^{(E_t - E_i)/k_B T}]} \tag{1.149}$$

假设电子的俘获截面与空穴的相等，即 $\sigma_n = \sigma_p = \sigma_0$，式(1.149)可简化为

$$U = v_{th}\sigma_0 N_t \frac{(pn - n_i^2)}{p + n + 2n_i \cosh\left[\dfrac{E_t - E_i}{k_B T}\right]} \tag{1.150}$$

考虑 n 型半导体。在小注入情形下，复合率可写为

$$U = \frac{p_n - p_{n0}}{\tau_r} \tag{1.151}$$

其中 τ_r 为复合寿命，由下式给出

$$\tau_r = \frac{1 + \left[\frac{2n_i}{n_{n0} + p_{n0}}\right]\cosh\left[\frac{E_t - E_i}{k_B T}\right]}{v_{th}\sigma_0 N_t} \tag{1.152}$$

由于 $n_{n0} \gg n_i$，复合寿命 τ_r 是 $(E_t - E_i)/k_B T$ 的缓变函数，即 τ_r 的最小值（等于 $1/v_{th}\sigma_0 N_t$）的范围是相当宽的。在 $n_{n0} = 10^{15}\text{cm}^{-3}$ 的情况下，即使复合中心能级 E_t 在禁带中线上下一个相当大的范围（达 $\pm 10k_B T$）内变化，τ_r 仍基本上为一个常数（$\approx 1/v_{th}\sigma_0 N_t$）。若多数载流子浓度更高，这个能量范围也变得更大。

到现在为止，我们考虑的都是过剩载流子注入半导体中，以致有 $pn > n_i^2$ 的情况，此时使系统恢复到平衡的是复合。另一方面，如果 $pn < n_i^2$，意味着载流子从半导体中抽出，为了恢复系统的平衡，产生-复合中心必然产生载流子。把 $n < n_i$ 及 $p < n_i$ 代入式(1.150)，即可得到产生率

$$G = -U = \frac{n_i}{\tau_g} \tag{1.153}$$

其中 τ_g 为产生寿命，由下式给出：

$$\tau_g = \frac{2\cosh\left[\frac{E_t - E_i}{k_B T}\right]}{v_{th}\sigma_0 N_t} \tag{1.154}$$

注意产生寿命强烈依赖于复合中心的能级位置，当复合中心能级位于禁带中线附近时，τ_g 取最小值。这是由于在低载流子浓度情形下，复合中心与导带和价带之间的跃迁速率趋向于相等时，这些能级起较大的作用。

由式(1.152)和(1.154)可得

$$\frac{\tau_g}{\tau_r} \approx 2\cosh\left[\frac{E_t - E_i}{k_B T}\right] \tag{1.155}$$

可见，当 $E_t \neq E_i$ 时，产生寿命比复合寿命大得多。式(1.155)对于器件设计具有重要意义。如果选择适当能级的复合中心，可使 $\tau_g \gg \tau_r$。因此，和 τ_r 有关的器件特性（如二极管的关断时间）可以独立改变，而不影响器件与 τ_g 有关的特性，如二极管的漏电流（见第二章）。

4. 俄歇复合

俄歇复合是电子和空穴直接复合、同时把释放出来的能量交给另一个电子或空穴（在 n 型半导体中是电子，在 p 型半导体中是空穴）的复合过程。俄歇复合寿命 τ_A 与多数载流子浓度的平方成反比，可表示为

$$\tau_A = \frac{1}{G_n n^2}(\text{n 型半导体}) \tag{1.156}$$

或

$$\tau_A = \frac{1}{G_p p^2}(\text{p 型半导体}) \tag{1.157}$$

G_n，G_p 称为俄歇系数，常用单位是 $[\text{cm}^6\text{s}^{-1}]$。对于硅，$Gn \approx 2.28 \times 10^{-31}$，$G_p \approx 9.9 \times 10^{-32}$。

尽管由于俄歇过程是一个三粒子过程，发生该过程的几率较小，但在高载流子浓度下，该过程可起重要作用。所以在重掺杂半导体中，俄歇复合是主要的复合机制。

5. 表面复合

由于晶体原子的周期排列在表面终止，在表面区引入了大量的局域能态或产生-复合中心。这些能态可以大大增加表面区域的复合率。因为表面复合对许多半导体器件的特性有很大的影响，所以了解表面复合过程是很重要的。

表面复合和前面讨论过的间接复合类似，是通过表面复合中心进行的。单位时间单位面积内载流子在表面复合的总数可用类似于式(1.143)的形式[并且利用了式(1.147)和(1.148)]表示为

$$U = \frac{v_{th}\sigma_p\sigma_n N_{st}(p_s n_s - n_i^2)}{\sigma_p(p_s + p_t) + \sigma_n(n_s + n_t)} \tag{1.158}$$

式中 n_s 和 p_s 分别为表面区中电子和空穴的浓度，N_{st} 为表面区域单位面积的复合中心数。在小注入的情形下，在 n_s 基本上等于体内的多数载流子浓度，因而 $n_s \gg p_s$ 以及 $n_s \gg n_t$ 的极限情况下，式(1.151)可简化为

$$U \approx v_{th}\sigma_p N_{st}(p_s - p_{n0}) \tag{1.159}$$

因乘积 $v_{th}\sigma_p N_{st}$ 的量纲是 cm/s，所以称之为小注入表面产生复合速度 S_r，即

$$S_r = v_{th}\sigma_p N_{st} \tag{1.160}$$

若假定表面的电子和空穴俘获截面相等($\sigma_n = \sigma_p = \sigma_s$)，则表面复合速度 S_r 和表面产生速度 S_g 可由式(1.158)推得：

$$S_r = \frac{v_{th}\sigma_s N_{st}}{1 + \left[\frac{2n_i}{n_{n0} + p_s}\right]\cosh\left[\frac{E_t - E_i}{k_B T}\right]} \tag{1.161}$$

$$S_g = \frac{v_{th}\sigma_s N_{st}}{2\cosh\left[\frac{E_t - E_i}{k_B T}\right]} \tag{1.162}$$

S_r 和 S_g 的表达式分别具有复合寿命和体产生寿命表达式的倒数形式。因此可预料，表面复合速度要比表面产生速度高，即 $S_r > S_g$。硅工艺控制的最新发展，已使 S_r 值低到80cm/s，S_g 则更低，为0.1cm/s。

表面复合使得表面处的载流子浓度降低，其浓度梯度产生扩散电流。结果，表面($x = 0$)附近的少数载流子(例如空穴)扩散电流由下式决定：

$$-D_p \left.\frac{dp_n}{dx}\right|_{x=0} = S_{pr}[p_n(x = 0) - p_{n0}] \tag{1.163}$$

式中，$S_{pr} = v_{th}\sigma_p N_{st}$ 为空穴的表面复合速度。

1.3.5 连续性方程和泊松方程

前面已经讨论了载流子的漂移、扩散、复合和产生。描述所有这些作用总体效应的基本方程为连续性方程。连续性方程基于粒子数守恒，即单位体积内电子增加的速率等于净流入的速率和净产生率($G_n - U_n$)之和：

$$\frac{\partial n}{\partial t} = \frac{1}{q}\nabla \cdot \boldsymbol{J}_n + G_n - U_n \tag{1.164}$$

式中，G_n 为外界因素作用(例如光子的光激发和强电场下的碰撞电离)下电子的产生率；U_n 为p型半导体中的电子复合率，在小注入(注入的载流子浓度远小于平衡多数载流子浓度)状

态下，U_n 可表示为$(n_p - n_{p0})/\tau_n$，式中 n_p 为电子(少数载流子)浓度，n_{p0}为热平衡电子浓度，τ_n 为电子寿命。若电子和空穴成对产生和复合，没有陷阱效应或其他效应，则 $\tau_n = \tau_p$(τ_p 为空穴寿命)。对于空穴也有类似的连续性方程：

$$\frac{\partial p}{\partial t} = -\frac{1}{q}\nabla \cdot \boldsymbol{J}_p + G_p - U_p \tag{1.165}$$

除连续性方程外，还必须满足泊松方程。这一方程由静电场强和电势梯度的关系 $\mathscr{E} = -\nabla\psi$和高斯定理的微分形式$\nabla \cdot \mathscr{E} = \rho/\varepsilon_s$得到，为

$$\nabla^2\psi = -\rho/\varepsilon_s \tag{1.166}$$

式中，ε_s是半导体的介电常数；$\rho = q(p - n + N_D^+ - N_A^-)$，即载流子浓度和电离杂质浓度的代数和。

1.3.6 强电场效应

1. 非线性的速度-电场关系

在讨论迁移率时，我们假定迁移率 μ(或两次碰撞间的时间间隔 τ)与电场 $\mathscr{E}$ 无关，从而漂移速度 v 与 $\mathscr{E}$ 呈线性关系。这一假定在弱电场下是成立的。实际上，只要载流子的平均能量偏离热平衡值$\frac{3}{2}k_BT$(T 晶格温度)很小，或者说只要漂移速度比热运动速度(室温下约为10^7cm/s)小得多，这个假定就是合理的。但随着 $\mathscr{E}$ 的不断增大，载流子从电场获得的能量越来越多，其平均能量(通常象对理想电子气那样用$\frac{3}{2}k_BT_e$ 表示，其中 T_e 称为热电子温度)将显著高于热平衡值，或者说漂移速度将接近热速度。在这种情形下，μ 将与 $\mathscr{E}$ 有关。这就是非线性 v-$\mathscr{E}$ 关系的起源。

图 1.16 为硅与砷化镓中电子及空穴的漂移速度和电场关系的实验结果。先考察硅的情

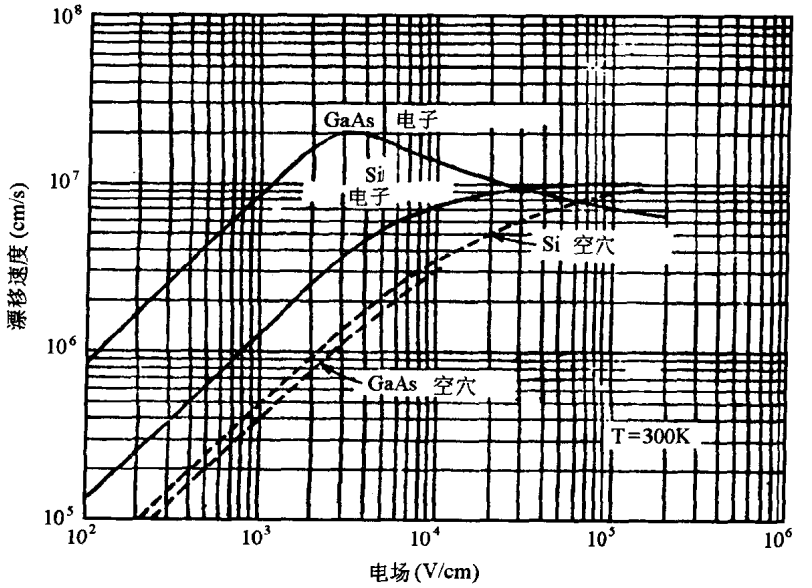


图 1.16 GaAs 与 Si 中漂移速度和电场的关系

况。显然，漂移速度与电场的关系起初是线性的，相应的迁移率是常数；随着电场进一步增加，漂移速度的增加减慢，当电场足够强时，漂移速度趋向于一个饱和值。饱和漂移速度起因于光学声子(能量为 $\hbar\omega_0$)对载流子能量损失速率的限制，对此我们进一步说明如下。当电场足够强时，载流子上有足够能量和光学声子相互作用，能量损失主要通过发射光学声子。平均说来，载流子通过发射光学声子损失能量的速率为

$$\left\langle \frac{dE}{dt} \right\rangle_1 = \frac{\hbar\omega_0}{\tau} \tag{1.167}$$

这里的时间 τ 就是载流子在光学声子散射下的平均自由时间。载流子从电场获得能量的平均速率为

$$\left\langle \frac{dE}{dt} \right\rangle_2 = qv\mathscr{E} \tag{1.168}$$

稳定时，$\left\langle \frac{dE}{dt} \right\rangle_1 = \left\langle \frac{dE}{dt} \right\rangle_2$ 和 $q\mathscr{E}\tau = mv$，故有

$$v = v_s = \sqrt{\hbar\omega_0/m} \tag{1.169}$$

v_s 表示载流子的饱和漂移速度。对硅中的电子，$\hbar\omega_0 \approx 0.05\text{eV}$，$m = m_\sigma \approx 0.26m_0$，($m_0$ 为自由电子质量)，v_s 约等于 $1.7\times10^7\text{cm/s}$。

n 型砷化镓的强场输运现象与硅很不相同。由图 1.16 可见，对 n 型 GaAs，漂移速度达到最大值后，将随着电场的进一步增加而减小。这种现象起因于 GaAs 的导带结构。在叙述多能谷半导体时曾经提到，GaAs 的导带底在 Γ 点($\boldsymbol{k}=0$)上；但在 L 点(沿<111>方向)上也存在着能谷，其能量约高 0.31eV。弱电场时，所有电子都留在低能谷(Γ 能谷)；电场增高，一些电子从电场获得足够能量转移到高能谷(L 能谷)；电场足够高时，所有电子都转移到 L 能谷。L 能谷的电子由于碰撞频繁(τ 小)和有效质量大，迁移率比 Γ 能谷的小得多。所以，当 k_BT_e 和两能谷的能量差可比时，大多数电子移入 L 能谷，电子的平均迁移率减小，即漂移速度减小，直至所有电子都转移到了 L 能谷为止。由于 n 型 GaAs 的漂移速度具有这种特点，因此 GaAs 材料被用来做微波转移电子器件(参见第六章)。

漂移速度与电场关系的实验结果也可以用经验公式近似表示，较为常用的经验公式是：

$$\text{Si：}\ v_n, v_p = \frac{v_s}{[1+(\mathscr{E}_0/\mathscr{E})^\gamma]^{1/\gamma}} \tag{1.170}$$

$$\text{GaAs：}\ v(\mathscr{E}) = \frac{\mu\mathscr{E} + v_s(\mathscr{E}/\mathscr{E}_0)^4}{1+(\mathscr{E}/\mathscr{E}_0)^4} \tag{1.171}$$

就 Si 而言，v_s 为 10^7cm/s；对于电子，γ 为 2，$\mathscr{E}_0$ 为 $7\times10^3\text{V/cm}$；对于空穴，γ 为 1，$\mathscr{E}_0$ 为 $2\times10^4\text{V/cm}$。对于 GaAs 中的电子，各参数的典型值是：

$$\mu = 8\,000\text{cm}^2/\text{V}\cdot\text{s},\quad v_s = 7.7\times10^6\text{cm/s},\quad \mathscr{E}_0 = 4\times10^3\text{V/cm}$$

2. *碰撞电离*

当半导体中的电场增至某值以上时，电子(或空穴)可获得足够高的动能与晶格碰撞，给出大部分动能打破一个价健，将一个价电子从价带电离到导带，因而产生一个电子-空穴对。这就是碰撞电离。在经典近似下，可以把碰撞电离过程想象为电子同静止的晶格原子之间的非弹性碰撞，电子损失的能量用来产生电子-空穴对，电离能(产生碰撞电离作用的电子必须具有的最低能量)E_i 应大于半导体的禁带宽度 E_g，这是因为碰撞时除能量守恒外，还必须满足动

量守恒条件。在上述经典模型的基础上经过简单分析，可以得到 $E_i \approx \frac{3}{2}E_g$。

根据上面的论述，只有动能(等于 $\frac{3}{2}k_B(T_e - T)$，T_e 是电子温度，T 是晶格温度)超过 E_i 的电子具有碰撞电离作用。对于玻耳兹曼分布，这部分电子所占比例为 e^{-E_i/k_BT_e}。电离率(一个电子或空穴在单位长度的路程上通过碰撞电离产生的电子-空穴对数目)，将正比于这一比例因子，

$$\alpha = Ae^{-E_i/k_BT_e} \tag{1.172}$$

A 为比例常数。电子从电场获得的动能为

$$\frac{3}{2}k_BT_e \approx q\mathscr{E}v_s\tau_E \tag{1.173}$$

τ_E 是电子的能量弛豫时间。由于碰撞电离是强电场作用，所以式(1.173)中用 T_e 代替了($T_e - T$)，用饱和漂移速度 v_s 代替了 $\mu\mathscr{E}$。利用式(1.173)可以将表示电离率 α 的式(1.172)改写为

$$\alpha = Ae^{-B/\mathscr{E}} \tag{1.174}$$

其中

$$B = \frac{3E_i}{2qv_s\tau_E} \tag{1.175}$$

对于一些重要的半导体，表1.3列出了电离率 $\alpha = Ae^{-B/\mathscr{E}}$ 中参数 A 和 B 的实验数据($\mathscr{E}$ 的单位用 V/cm)。

表 1.3　$\alpha = Ae^{-B/\mathscr{E}}$ 中的参数 A 和 B

半导体材料	电子		空穴	
	$A(cm^{-1})$	$B(V/cm)$	$A(cm^{-1})$	$B(V/cm)$
Ge	1.61×10^7	1.60×10^6	1.04×10^{-7}	1.28×10^6
Si	2.24×10^6	1.61×10^6	1.01×10^6	2.11×10^6
GaAs	1.05×10^6	1.52×10^6	1.70×10^6	1.85×10^6
GaP	2.26×10^6	3.33×10^6	2.26×10^6	3.33×10^6
InP	3.48×10^6	2.76×10^6	2.64×10^6	2.40×10^6

实验发现，只有少数半导体的电子电离率 α_n 和空穴电离率 α_p 是一样的。一般情形下，$\alpha_n \neq \alpha_p$，应有

$$\alpha_n = A_ne^{-B_n/\mathscr{E}} \tag{1.176}$$

$$\alpha_p = A_pe^{-B_p/\mathscr{E}} \tag{1.177}$$

碰撞电离是强电场引起的重要的产生过程。由于电子和空穴成对产生，产生率既可用单位体积每秒产生的电子数 dn/dt 表示，也可以用相应的空穴数(或电子-空穴对数)表示。同时，碰撞电离是电子和空穴都参与的过程，产生率应该是它们的贡献之和。由于电子的碰撞电离作用，电子的增加率为 $(dn/dt) = (dn/dx)(dx/dt) = n\alpha_nv_n$，空穴的增加率显然也是 $n\alpha_nv_n$；同理，由于空穴的碰撞电离作用，电子(或空穴)的增加率是 $p\alpha_pv_p$。将二者的作用相加，即得碰撞电离的产生率为

$$G_I = n\alpha_nv_n + p\alpha_pv_p$$

$$= A_n e^{-B_n/\mathscr{E}}(nv_n) + A_p e^{-B_p/\mathscr{E}}(pv_p) \tag{1.178}$$

n 和 v_n 分别为电子的浓度和运动速度，p 和 v_p 分别为空穴的浓度和运动速度。若 $\alpha_n = \alpha_p = \alpha$，则式(1.178)简化为

$$G_I = \alpha(nv_n + pv_p) \tag{1.179}$$

1.3.7 非稳态输运效应：速度过冲

前面我们实际上只讨论了稳态输运问题。与之相应，描述半导体输运性质的基本方程(漂移-扩散方程和泊松方程)只适合于常规器件，即只适合于大尺寸的器件(如沟道长度大于亚微米)和低于一定工作频率(例如 30GHz)的器件。目前，随着微细加工技术的发展，半导体器件的尺寸越来越小，已经能做到 0.1μm(深亚微米)的量级。对于深亚微米器件，非稳态效应将出现，如产生载流子速度过冲。

在前面的讨论中一直认为，对应于某一确定的电场存在恒定不变的漂移速度。显然，这是一个稳态输运的概念。在电场对于载流子的加速时间小于能量弛豫时间的尺度内，漂移速度将随时间变化，即不能再把漂移速度看作电场的函数。图 1.17 表示在时间阶梯电场作用下，Si 和 GaAs 中电子漂移速度的瞬态特征。由图可见，在加上电场以后十分之几皮秒(ps)时间内，漂移速度急剧上升，并达到一极大值，它比稳态值大若干倍。漂移速度达到极大值以后就逐渐减小，并趋于稳态值。这种现象称为速度过冲。

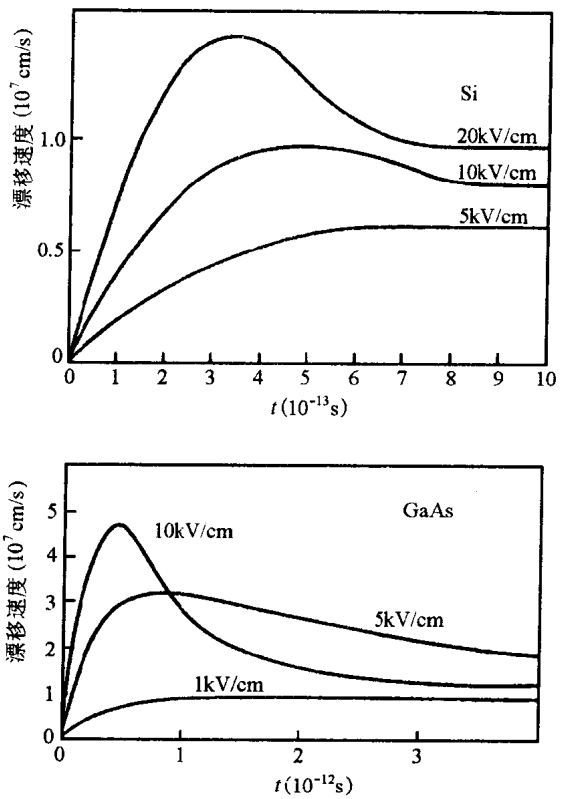


图 1.17 Si 和 GaAs 中电子在时间阶梯电场作用下漂移速度随时间的变化

速度过冲现象可定性解释如下。在加上电场以后，电子被加速，电子能量也逐渐增加。由于能量弛豫时间比动量弛豫时间长得多，在动量弛豫时间的时间尺度内，电子能量要低于将要达到的稳定能量。电子能量较低通常意味着受到散射作用较小(即动量弛豫时间较长)，从而在一个短暂的时间内可以达到一个高于稳态值的漂移速度。然后，随着时间的推移，电子能量将逐渐上升到一稳定值，电子漂移速度因散射作用逐渐下降到一稳定值。在 Si 和 Ge 中，声子散射速率强烈依赖于载流子能量，上述定性解释是直截了当的。而在 GaAs 和其他某些Ⅲ-Ⅴ族化合物中，Γ 能谷中主要的散射过程由极性光学声子引起，电子能量足够高时这些散射过程的速率近似与能量无关，所以只有在高于发生电子转移效应的阈值电场(GaAs 中≥3.2kv/cm，InP 中≥11kv/cm)下才出现速度过冲。同时，在这些材料中，代替能量弛豫时间的应是能量上升的时间加上从低能谷向高能谷转移的时间，速度过冲现象尤其显著。

速度过冲现象对于深亚微米器件的工作速度是十分重要的。为此，把电子漂移速度作为从源算起的距离 $\left(x(t)=\int v(t')\mathrm{d}t'\right)$ 的函数是更直截了当的。作为简单估计，我们可以只是把图 1.17 中的时间坐标乘以饱和速度 $v_s\approx10^7\mathrm{cm/s}$，就把漂移速度与时间的关系变换为漂移速度与距离的关系。从图 1.17 的数据不难看出，Si 中速度过冲的特征长度(即电子以高于稳态漂移速度行进的距离)为几百埃的量级；而在 GaAs 中，速度过冲的特征长度可以大到 1 微米。在 GaAs 晶体管的设计中，速度过冲的蒙特-卡罗模拟是十分有用的工具。

1.4 半导体的光学性质

1.4.1 引言

正像晶格振动能够引起电子在不同状态之间的跃迁一样，在光的作用下，半导体中的电子也可以在不同状态之间跃迁并引起光的吸收或发射。这种跃迁可以发生在不同能带的状态之间，或同一能带内的各个状态之间，也可以发生在分立能级和能带之间。半导体器件中最重要的光电子相互作用是能带间的跃迁。对于能带间跃迁，在光子吸收的过程中，一个光子把价带中的一个电子散射到导带内；在相反的过程中，导带电子同价带一个空穴复合而产生一个光子(这种情况称为辐射跃迁；也有不辐射出光的情况，此时由于碰撞而失去能量，这就是无辐射跃迁)。这两个过程显然对光探测和光发射的器件是十分重要的。本节叙述的半导体光学性质将为讨论光电子器件提供一定的物理基础。

电子在跃迁过程中不但要满足能量守恒，而且要满足动量守恒。在光跃迁中，它们可分别表示为：

(1) 能量守恒　在吸收和发射过程中，对于电子的终态能量 E_f 和初态能量 E_i 有

吸收　　$E_f=E_i+h\nu$

发射　　$E_f=E_i-h\nu$

$h\nu$ 是光子能量，其中 h 是普朗克常数，ν 是光子频率。由于导带状态和价带态的最小能量差是禁带宽度 E_g，光子能量必须高于 E_g 才能发生带间吸收。

(2) 动量守恒($\boldsymbol{k}$ 守恒)　光子动量之值为 $\hbar k_L=h/\lambda$，其中 k_L 是光子波矢，$\hbar=h/2\pi$ 是约化普朗克常数。由于能量 1eV 的光子对应波长 $\lambda=1.24\mu\mathrm{m}$，$k_L$ 为 $10^4\mathrm{cm}^{-1}$的数量级，与晶体电子的 k 值(k 在布里渊区的尺度内，即 $k\sim1/a$，a 是晶格常数)相比是微不足道的。所以，在

光跃迁过程中电子波矢 $\boldsymbol{k}$ 基本不变，这就是所谓的 $\boldsymbol{k}$ 选择定则。

由于 $\boldsymbol{k}$ 守恒的限制，价带边和导带边同在 $\boldsymbol{k}=0$ 的半导体（直接禁带半导体）中，只有竖直跃迁是允许的。这种光跃迁不需要声子参与，因而是很强的。而在 Si 和 Ge 等间接禁带半导体中，需要晶格振动以满足 $\boldsymbol{k}$ 守恒，带边附近的光跃迁很弱。

因为竖直跃迁中电子和空穴的 $\boldsymbol{k}$ 值相同，对于光子能量 $\hbar\omega$（$\omega=2\pi\nu$ 是光子的角频率）大于禁带宽度 E_g 的情形，有

$$\hbar\omega = E_g + \frac{\hbar^2k^2}{2}\left[\frac{1}{m_n}+\frac{1}{m_p}\right] = E_g + \frac{\hbar^2k^2}{2m_r} \tag{1.180}$$

m_r 是电子-空穴的折合质量。这样，在竖直跃迁中，相关的态密度函数中的有效质量应当为折合质量。这种态密度称为联合态密度。

1.4.2 辐射跃迁和光吸收

在包括半导体在内的原子系统中，光子和电子之间的相互作用有三种基本过程：吸收、自发发射和受激发射。当一个能量 $h\nu_{12}=E_2-E_1$ 的光子入射到这个系统上时，一个处于低能态 E_1 的粒子就要吸收这个光子，跃迁到高能态 E_2，这个过程就是吸收过程，如图 1.18(a)所示。粒子的高能态是不稳定的，在一段时间内，如果没有外界激发，它又自动回到低能态 E_1，并发射一个能量 $h\nu_{12}$的光子，这种过程称为自发发射过程，如图 1.18(b)所示。在能量 $h\nu_{12}$的光子作用下，已经处在高能态 E_2 的粒子将受到激发，跃迁回到低能态 E_1，并发射出一个能量 $h\nu_{12}$的光子，这种过程称为受激发射过程，如图 1.18(c)所示。在自发发射过程中，产生的光子在传播方向、相位和偏振状态上是随机的，彼此无关，出射光为非相干光，半导体发光二极管就是利用这种自发发射效应出光。在受激发射过程中，产生的光子和入射光子具有相同的频率、传播方向、偏振状态和相位，即入射光得到放大，出射光是相干光，半导体激光器就是利用这个原理制成的。光电探测器和太阳电池则是利用了光吸收过程。

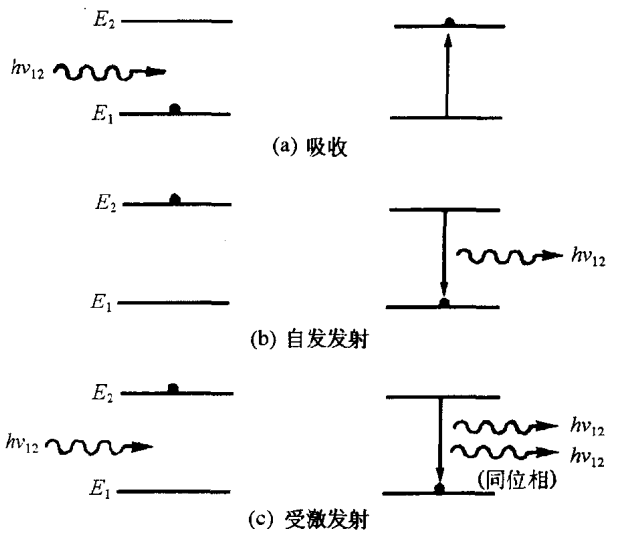


图 1.18　两能级之间的基本跃迁过程（黑点表示粒子的状态）

1. 自发发射

假定系统中处在高能态 E_2 的粒子的浓度为 N_2，则自发向低能态 E_1 跃迁而减少的速率可表示为

$$-\frac{dN_2}{dt} = A_{21}N_2 \tag{1.181}$$

A_{21}是比例系数，表示单位时间内每个 E_2 能态粒子向能态 E_1 跃迁的几率，即自发跃迁速率。

2. 受激发射

在光子能量 $\hbar\omega = E_2 - E_1$ 的光波作用下，高能态 E_2 的粒子减少的速率可表示为

$$-\frac{dN_2}{dt} = B_{21}\rho(\hbar\omega)N_2 \tag{1.182}$$

比例系数 B_{21}称为受激发射(吸收)系数，$\rho(\hbar\omega)$是光子密度，$B_{21}\rho(\hbar\omega)$表示光作用下单位时间内每个 E_2 能态粒子发生 2→1 的跃迁并发射能量为 $\hbar\omega$ 的光子的几率，即受激发射速率。

3. 受激吸收

由于吸收能量 $\hbar\omega = E_2 - E_1$ 的光子，粒子发生 1→2 的受激跃迁，低能态 E_1 的粒子减少的速率可表示为

$$-\frac{dN_1}{dt} = B_{21}\rho(\hbar\omega)N_1 \tag{1.183}$$

N_1 表示系统中处在低能态 E_1 的粒子的浓度。

热平衡的系统中上述三个过程虽然在不断进行，但在宏观上吸收和发射应互相抵消，即

$$[A_{21} + B_{21}\rho(\hbar\omega)]N_2 = B_{12}\rho(\hbar\omega)N_1 \tag{1.184}$$

可以解出

$$\rho(\hbar\omega) = \frac{A_{21}/B_{21}}{\frac{B_{12}}{B_{21}}\frac{N_1}{N_2} - 1} \tag{1.185}$$

当 $E_2 - E_1 > 3k_BT$ 时，粒子数之比由玻耳兹曼因子决定：

$$\frac{N_2}{N_1} = e^{-\hbar\omega/k_BT} \tag{1.186}$$

另一方面，$\rho(\hbar\omega)$也可以理论计算如下。对于实际空间体积，波数 k 和 $k + dk$ 范围内的光子状态(光子模式)数为

$$2 \times \frac{4\pi k^2 dk}{(2\pi)^3}$$

因子 2 表示光子有两种彼此独立(相互正交)的偏振状态。在折射率为 $\bar{n}$ 的介质中，$k = \frac{2\pi}{\lambda} = \frac{\omega\bar{n}}{c}$，$c$ 是真空中的光速，故上式又可表示成

$$\frac{\omega^2\bar{n}^3}{\pi^2\hbar c^3}$$

根据玻定-爱因斯坦分布，平衡时能差为 $\hbar\omega$ 的光子模所包含的光子数 $N(\hbar\omega)$ 为 $1/[\exp(\hbar\omega/k_BT) - 1]$，可把 $\rho(\hbar\omega)$一般地表示为

$$\rho(\hbar\omega) = \frac{\omega^2\bar{n}^3}{\pi^2\hbar c^3} \cdot \frac{1}{e^{\hbar\omega/k_BT} - 1} \tag{1.187}$$

将式(1.187)代入式(1.185),并注意所得关系式对任何温度下都应当成立,可得

$$B_{12} = B_{21} \tag{1.188}$$

$$A_{21} = \frac{\omega^2 \bar{n}^3}{\pi^2 \hbar c^3} B_{12} \tag{1.189}$$

式(1.188)表明两个量子态之间互为逆过程的跃迁几率相等,这正是从细致平衡原理可以预期的结果。式(1.189)表明,粒子的自发寿命 τ_0(等于 $1/A_{21}$,A_{21}代表 E_2 能态的粒子每秒自发向能态 E_1 跃迁的平均次数)可通过计算量子跃迁几率得到,为(可参阅叶良修《半导体物理学》§15.1):

$$A_{21} = \frac{q^2 \bar{n} \hbar \omega}{\pi m_0^2 \varepsilon_0 c^3 \hbar^2} |\mathbf{e} \cdot \boldsymbol{p}_{21}|^2 \tag{1.190}$$

m_0 是自由电子质量;ε_0 是真空介电常数;$\mathbf{e}$ 为偏振方向的单位矢量;$\boldsymbol{p}_{21}$为状态 2 和 1 之间的动量矩阵元。

1.4.3 直接禁带半导体中带间跃迁的自发发射和载流子寿命

上面用两能级模型介绍了光发射和光吸收的基本理论,现在将这一理论用于直接禁带半导体的自发发射。为此,应当考虑下述两个方面:(1) 上、下能态分别对应导带态和价带态,这些态准连续分布并且遵从费米-狄拉克统计;(2) 对于带间跃迁,遵从 $\boldsymbol{k}$ 守恒定则,即对应每一个 $\boldsymbol{k}_c$ 的导带态有一个并且只有一个相应的价带态 $\boldsymbol{k}_v = \boldsymbol{k}_c$。所以单位体积中参与自发发射的状态数是 $\rho(k)\mathrm{d}k$,$\rho(k)$是 k 空间的能态密度。基于这些考虑,自发发射的速率为

$$R_{sp} = \int A \rho(k) f_n f_p \mathrm{d}k \tag{1.191}$$

其中,f_n 为导带态 E_n 上电子的占有几率,f_p 为价带能级 E_p 上空穴的占有几率,可用各自的准费米能级 E_{Fn}和 E_{Fp}分别表示如下:

$$f_n = [1 + \mathrm{e}^{(E_n - E_{Fn})/k_B T}]^{-1} \tag{1.192}$$

$$f_p = [1 + \mathrm{e}^{(E_{Fp} - E_p)/k_B T}]^{-1} \tag{1.193}$$

A(或 $1/\tau_0$)由式(1.190)得到为

$$A = \frac{q^2 \bar{n} \hbar \omega}{\pi m_0^2 \varepsilon_0 c^3 \hbar^2} |\boldsymbol{e} \cdot \boldsymbol{p}_{cv}|^2 \tag{1.194}$$

$\boldsymbol{p}_{cv}$为导带和价带之间的动量矩阵元。对于绝大多数半导体,$2p_{cv}^2/m_0 \approx 20\mathrm{eV}$。在 τ_0 的定义中,假定了电子可以找到一个与之复合的空穴。如果找到这种空穴的几率很小,则辐射的时间就会很长。对于 GaAs 这样的材料,τ_0 之值约为 1ns,但对于间接禁带材料,复合时间可长达 $1\mu\mathrm{s}$。

在 k 值附近,导带电子能量 E_n 和价带空穴能量 E_p 可表示为

$$E_n = E_c + \frac{\hbar^2 k^2}{2m_n} \tag{1.195}$$

$$E_p = E_v + \frac{\hbar^2 k^2}{2m_p} \tag{1.196}$$

m_n 和 m_p 分别是电子和空穴的有效质量,于是跃迁能量为

$$\hbar\omega = E_n - E_p = E_g + \frac{\hbar^2 k^2}{2m_r} \tag{1.197}$$

其中 $1/m_r = 1/m_n + 1/m_p$ 为折合质量。对于半导体单位体积，k 值落入 dk 的球壳内的能级对数为

$$\rho(k)dk = 2 \times \frac{4\pi k^2}{(2\pi)^3}dk \tag{1.198}$$

所以，

$$\rho(k) = k^2/\pi^2 \tag{1.199}$$

下面改用能量 $\hbar\omega$ 作为积分变量。为此，利用式(1.197)得到

$$\rho(k)dk = \frac{\sqrt{2}\,m_r^{3/2}(\hbar\omega - E_g)^{1/2}}{\pi^2\hbar^3}d(\hbar\omega) = g_{cv}(\hbar\omega)d(\mu\omega) \tag{1.200}$$

$g_{cv}(\hbar\omega)$称为联合态密度。将式(1.194)及(1.200)代入式(1.191)，可得

$$R_{sp} = \int \frac{1}{\tau_0}d(\hbar\omega)g_{cv}(\hbar\omega)f_n f_p \tag{1.201}$$

对于带隙附近的跃迁，我们假定 τ_0 与 ω 无关，于是上式简化为

$$R_{sp} = \frac{1}{\tau_0}\int d(\hbar\omega)g_{cv}(\hbar\omega)f_n(E_n)f_p(E_p) \tag{1.202}$$

除了自发发射速率以外，实际感兴趣的另一个物理量是载流子的辐射寿命。对于小注入情形下的非简并半导体，费米分布可用玻耳兹曼分布近似，有

$$f_n f_p \approx e^{-[E_c - E_{Fn}]/k_B T} \cdot e^{-(E_{Fp} - E_v)/k_B T} \cdot e^{-(\hbar\omega - E_g)/k_B T} \tag{1.203}$$

将式(1.203)用于方程(1.202)，得到

$$R_{sp} = \frac{1}{2\tau_0}\left(\frac{2\pi\hbar^2 m_r}{k_B T}\right)^{3/2} np \tag{1.204}$$

若把总电荷表示为热平衡电荷加过剩电荷($n = n_0 + \Delta_n$, $p = p_0 + \Delta_p$)，并注意到热平衡时复合速率和产生速率相等，则不难证明，净复合率是 $\Delta n/\tau_r$，其中

$$\frac{1}{\tau_r} = \frac{1}{2\tau_0}\left(\frac{2\pi\hbar^2 m_r}{k_B T}\right)^{3/2}(n_0 + p_0) \tag{1.205}$$

τ_r 为过剩载流子的辐射寿命。τ_r 远大于 τ_0，这是因为小注入时电子找到与之复合的空穴的几率很小。

另一种重要情形是简并半导体。设样品为重掺杂 n 型，且 $n \gg p$，这时可以假定 $f_n = 1$ 和 f_p 服从玻耳兹曼分布，由方程(1.202)可以求得自发跃迁速率为

$$R_{sp} = \frac{1}{\tau_0}\left(\frac{m_r}{m_p}\right)^{3/2} \tag{1.206}$$

空穴寿命 τ_r 为

$$\frac{1}{\tau_r} = \frac{1}{\tau_0}\left(\frac{m_r}{m_p}\right)^{3/2} \tag{1.207}$$

同样，对简并 p 型半导体，电子寿命为

$$\frac{1}{\tau_r} = \frac{1}{\tau_0}\left(\frac{m_r}{m_n}\right)^{3/2} \tag{1.208}$$

综上所述，对于每个过剩载流子的辐射复合时间 τ_r 通常可以写成

$$\tau_r = \Delta n/R_{sp} \tag{1.209}$$

在简并半导体或大注入的情形下，$\tau_r \approx \tau_0$；一般情形下，τ_r 同 R_{sp}的关系强烈依赖于载流子浓度。

1.4.4 直接禁带半导体中的光吸收和光增益

光是一种电磁波，在包括半导体在内的介质中传播时，具有由电场矢量$\mathscr{E}$所代表的形式：

$$\mathscr{E} = \mathscr{E}_0 \, e^{i\omega\left(\frac{\bar{n}x}{c} - t\right)} \cdot e^{-\alpha x/2} \tag{1.210}$$

x 是传播方向，ω 是频率，$\bar{n}$ 是折射率(实部)，α 是介质的吸收系数。若 α 为零，电磁波无衰减传播，速度为 $c/\bar{n}$(c 是真空中的光速)；但对于 α 不为零，光子流密度 $I(\sim \mathscr{E}^* \mathscr{E})$下降为

$$I(x) = I(o)e^{-\alpha x} \tag{1.211}$$

对于直接禁带半导体的带间吸收，α 可表示为(详细推导可参阅叶良修《半导体物理学》§13.1)

$$\alpha = \frac{\pi q^2 \hbar}{\bar{n} c m_0^2 \hbar\omega \varepsilon_0} |p_{cv}|^2 g_{cv}(\hbar\omega) \tag{1.212}$$

$g_{cv}(\hbar\omega)$为联合态密度，其表示式可由(1.200)给出，为

$$g_{cv}(\hbar\omega) = \frac{\sqrt{2} m_r^{3/2}}{\pi^2 \hbar^3} (\hbar\omega - E_g)^{1/2} \tag{1.213}$$

p_{cv}为动量矩阵元，对于非偏振光，平均的矩阵元为$(2/3)p_{cv}^2$。计算表明 GaAs 对非偏振光的吸收系数可表示为

$$\alpha(\hbar\omega) \approx 5.6 \times 10^4 \frac{(\hbar\omega - E_g)^{1/2}}{\hbar\omega} \tag{1.214}$$

其中 $\hbar\omega$ 和 E_g 的单位用 eV。对于任何其他直接禁带半导体，前面的数值因子可以通过折合质量比的 3/2 次方按比例得到。

上面处理的是一种理想情况：导带是完全空的，价带则完全被电子填满。所以，只存在同向上跃迁相联系的光吸收，而不存在同向下跃迁相联系的光发射。现在，设想电子注入导带，而空穴注入价带。这时，一方面紧靠带边的吸收跃迁将被减弱；另一方面，由导带向价带的发射跃迁的几率增加了，这可导致适当频率的光的增益。

下面我们将定义光的增益系数 G 来代替吸收系数α：

$$dI = GI dx \tag{1.215}$$

入射光强 I 随透射深度 x 的变化可表示为

$$I = I_0 e^{Gx} \tag{1.216}$$

G 和 α 有相同的量纲(cm^{-1})。在导带和价带被电子和空穴分别部分占据的一般情形下，对增益 G 的贡献将来自两种类型的跃迁：由导带向价带的发射跃迁所贡献的 G_1 和由价带向导带的吸收跃迁所贡献的 G_2。前者引起光的增益，具有正值；后者引起光的吸收，具有负值。

考虑掺杂浓度很低的直接带隙半导体，假设跃迁中保持 k 守恒。当导带和价带分别被电子和空穴部分地占据时，吸收跃迁对增益的贡献为

$$G_2 = -\alpha(\hbar\omega)[(1 - f_n)(1 - f_p)] \tag{1.217}$$

其中 $\alpha(\hbar\omega)$表示价带全满、导带全空时对能量 $\hbar\omega$ 的光子的本征吸收系数，$(1 - f_n)$为导带能级被空穴占据的几率，$(1 - f_p)$为价带能级被电子占据的几率。类似地，发射跃迁的贡献的 G_1

可表示为

$$G_1 = \alpha(\hbar\omega) f_n f_p \tag{1.218}$$

通过向导带和价带注入大量的过剩载流子(它们首先将填充带边附近的状态),在适当高于 E_g 的光频下增益超过吸收而导致净增益是可能的。根据上述,增益系数为

$$\begin{aligned} G &= \alpha(\hbar\omega)[f_n f_p - (1 - f_n)(1 - f_p)] \\ &= \alpha(\hbar\omega)[f_n + f_p - 1] \end{aligned} \tag{1.219}$$

若 $f_n = 0$ 和 $f_p = 0$,即导带没有电子和价带没有空穴,增益系数简单地为 $-\alpha(\hbar\omega)$。在下述条件下,增益系数为正:

$$f_n + f_p - 1 > 0 \tag{1.220}$$

式(1.220)称为粒子数反转条件。这时,$G > 0$,通过光发射产生的附加光子使光强增加。光增益是半导体激光器的基础。在非偏振光的情形下,GaAs 的增益为

$$G(\hbar\omega) \approx 5.6 \times 10^4 \frac{(\hbar\omega - E_g)^{1/2}}{\hbar\omega}[f_n(E_n) + f_p(E_p) - 1] \tag{1.221}$$

对于折合质量 $m_r(\mathrm{A})$ 的半导体,前面的数值因子可通过乘以 $(m_r(\mathrm{A})/m_r(\mathrm{GaAs}))^{3/2}$ 得到。

第二章　p-n 结

现代电子技术中广泛使用 p-n 结作为整流、开关及其他用途的器件，半导体微波器件及光电器件的基本结构也是 p-n 结，同时，p-n 结也是双极型晶体管、可控硅整流器和场效应晶体管的基本组成部分，所以 p-n 结在对半导体器件的理解方面起着十分重要的作用。本章讨论由同种半导体材料的 p 型区和 n 型区构成的 p-n 结的性质。

p-n 结的最重要的性质是它的整流效应，即它只允许电流沿一个方向通过。图 2.1 是一个典型 p-n 结的伏-安特性，当在 p-n 结加一正向偏置时(偏置的极性在 2.2 节中说明)，电流随着偏压增加而迅速增大。但是当加"反向偏置"时，开始阶段几乎没有电流，反向偏压增加，电流也一直很小，当电压达到临界点时电流急剧增加。这种电流突然增大的现象称为结的击穿，所加的正向电压通常小于 1V，但反向击穿电压可以从几伏到几千伏，这和掺杂浓度以及器件的其他参数有关。

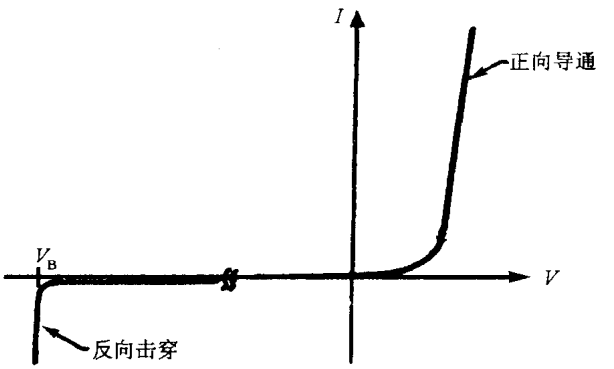


图 2.1　典型 p-n 结的伏-安特性

2.1　热平衡状态

一块 p 型半导体和一块 n 型半导体紧密接触就是一个 p-n 结。形成 p-n 结的方法很多，例如用一块 n 型半导体，可以通过高温扩散或离子注入从表面掺入 p 型杂质，形成一层 p 型半导体，也可以在它上面生长一层 p 型半导体。后一类方法称为外延(epitaxy)，有汽相、液相和分子束外延。为简单起见，本节考虑 p 区和 n 区都为均匀掺杂的情形，即突变 p-n 结。

p-n 结中有很大的载流子浓度梯度，它将引起载流子的相互扩散，p 区空穴向 n 区扩散，n 区电子向 p 区扩散。这样，在交界面附近，p 区留下了不能移动的带负电荷的电离受主，n 区留下了不能移动的带正电荷的电离施主，即形成所谓空间电荷区。空间电荷区产生电场，其方向从正电荷指向负电荷。空间电荷区也称势垒区。

空间电荷区电场的方向和载流子扩散流动的方向相反。在热平衡下，也就是在没有外界激发并给定温度的稳定条件下，通过结的净电流为零。这样，无论电子或空穴，由电场引起的

漂移电流和由浓度梯度引起的扩散电流恰好抵消。可以证明,在热平衡下,整个 p-n 结样品的费米能级 E_F 为一常数。

图 2.2 表示热平衡时突变 p-n 结的能带图。在图中,$x \leqslant -x_p$ 的区域为 p 型中性区;$x \geqslant x_n$ 的区域为 n 型中性区;$-x_p \leqslant x \leqslant x_n$ 的区域为空间电荷区(耗尽区)。

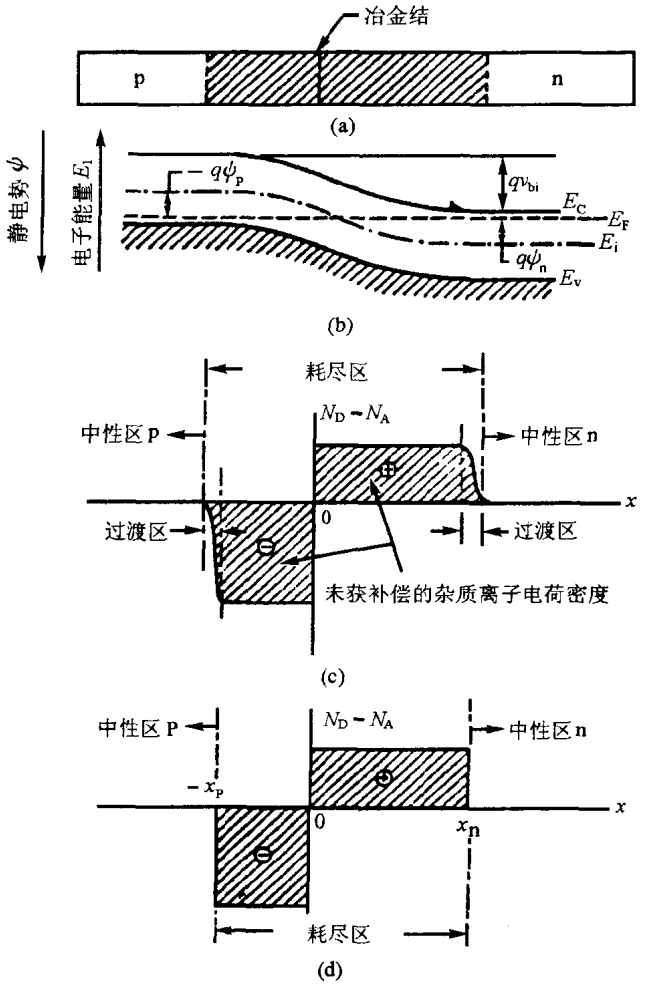


图 2.2 (a) 掺杂浓度在冶金结突变时的 p-n 结;(b) 突变结在热平衡下的能带图;(c) 空间电荷分布;(d) 空间电荷分布的矩形近似

在半导体中,导带底 E_c 相当于电子势能。但势能可以差一常数,因此可以用能带图上平行于 E_c 的任何一条曲线来表示电子势能。下面我们用本征费米能级 E_i 来表示电子势能 $-q\psi$,ψ 是相应的静电势。这样,电场 $\mathscr{E} = -\mathrm{d}\psi/\mathrm{d}x$ 可表示为

$$\mathscr{E} = \frac{1}{q}\frac{\mathrm{d}E_i(x)}{\mathrm{d}x} \tag{2.1}$$

空穴电流密度为

$$J_P = q\mu_P p\mathscr{E} - qD_P\frac{\mathrm{d}p}{\mathrm{d}x}$$

$$= q\mu_P p\left(\frac{1}{q}\frac{dE_i(x)}{dx}\right) - k_B T\mu_P \frac{dp}{dx} \tag{2.2}$$

μ_P 和 D_P 分别为空穴的迁移率和扩散系数。上式中已用了爱因斯坦关系 $D_P = \frac{k_B T}{q}\mu_P$。将空穴浓度的表示式

$$p = n_i \exp\left(\frac{E_i - E_F}{k_B T}\right) \tag{2.3}$$

和它的导数

$$\frac{dp}{dx} = \frac{p}{k_B T}\left[\frac{dE_i}{dx} - \frac{dE_F}{dx}\right] \tag{2.4}$$

代入式(2.2),可得到空穴密度为

$$J_P = \mu_P p \frac{dE_F}{dx} \tag{2.5}$$

同理,电子电流密度为

$$J_n = \mu_n p \frac{dE_F}{dx} \tag{2.6}$$

热平衡时,既无空穴电流,又无电子电流,所以有

$$\frac{dE_F}{dx} = 0 \tag{2.7}$$

即费米能级 E_F 在整个样品内为常数(与 x 无关),如图 2.2 所示。

由图 2.2 可见,p 型中性区($x \leqslant -x_P$)内相对费米能级的静电势可表示为

$$\psi_p = \frac{(E_i - E_F)_{x \leqslant -x_P}}{-q} \tag{2.8}$$

利用 $p = n_i \exp\left(\frac{E_i - E_F}{k_B T}\right)$ 和 $p \simeq N_A$(受主浓度),得到

$$\psi_P = -\frac{k_B T}{q}\ln\left(\frac{N_A}{n_i}\right) \tag{2.9}$$

同理,n 型中性区($x \geqslant x_n$)内的静电势可表示为

$$\psi_n = \frac{(E_i - E_F)_{x \geqslant x_n}}{-q} \tag{2.10}$$

利用 $n = n_i \exp\left(\frac{E_F - E_i}{k_B T}\right)$ 和 $n \simeq N_D$(施主浓度),得到

$$\psi_n = \frac{k_B T}{q}\ln\left(\frac{N_D}{n_i}\right) \tag{2.11}$$

热平衡时,n 型中性区和 p 型中性区之间总的电势变化为

$$V_{bi} = \psi_n - \psi_P = \frac{k_B T}{q}\ln\frac{N_A N_D}{n_i^2} \tag{2.12}$$

V_{bi}称为内建电势,它实际上是组成 p-n 结的 n 型材料和 p 型材料之间的接触电势差。按照多种材料串联接触的电势差在同一温度下只决定于第一种材料和最后一种材料而和中间任何一种材料无关这一性质,不难理解,用普通电表测不出 p-n 的内建电势;无论画电路图或列电路方程都不应当考虑这个电势差。但内建电势影响载流子分布,在研究半导体器件中的物理过

程时是十分重要的。

2.2 耗尽区和耗尽层电容

本节考虑最重要的两类 p-n 结：突变结和线性缓变结，讨论耗尽区内的电场分布和电势分布，以及它的电容性质。

2.2.1 热平衡情形

p-n 结内电势 $\psi(x)$和电荷密度分布 $\rho(x)$之间的关系由泊松方程决定：

$$\frac{d^2\psi}{dx^2} = -\frac{\rho(x)}{\varepsilon_s}$$

$$= -\frac{q}{\varepsilon_s}(N_D - N_A + p - n) \tag{2.13}$$

ε_s是半导体的介电常数，可以表示为 $\varepsilon_s = \varepsilon_r\varepsilon_o$，$\varepsilon_r$ 为相对介电常数，是一个无量纲的数，对于 Si，$\varepsilon_r \simeq 12$；ε_0 为真空介电常数，$\varepsilon_0 \simeq 8.85\times10^{-12}$F/m，在 p-n 结的空间电荷区内，载流子数目很小，作为一级近似，令 $n = p = o$，这是所谓的耗尽近似，空间电荷区也通常称为耗尽区，其空间电荷全部为电离杂质电荷。

1. 突变结

热平衡时的空间电荷分布和电场分布如图 2.3 所示，泊松方程为

$$\frac{d^2\psi}{dx^2} = \begin{cases} qN_A/\varepsilon_s & (-x_p \leqslant x \leqslant 0), \\ -qN_D/\varepsilon_s & (0 \leqslant x \leqslant x_n)。 \end{cases} \tag{2.14}$$

将上式积分一次，并利用耗尽区边界及其外面区域内($x \leqslant -x_P$, $x \geqslant x_n$)电场为零的条件，得到耗尽区内的电场分布为

$$\mathscr{E}(x) = -d\psi/dx$$

$$= \begin{cases} qN_A(x + x_P)/\varepsilon_s & (-x_p \leqslant x \leqslant 0), \\ qN_D(x - x_n)/\varepsilon_s & (0 < x \leqslant x_n)。 \end{cases} \tag{2.15}$$

n 区一侧正电荷(施主离子)发出的电力线都终止于 p 区一侧的负电荷(受主离子)，因此在界面($x = 0$)处电场强度最大，数值为

$$\mathscr{E}_m = \frac{qN_Ax_P}{\varepsilon_s} = \frac{qN_Dx_n}{\varepsilon_s} \tag{2.16}$$

半导体内总体呈中性的条件要求

$$N_Ax_P = N_Dx_n \tag{2.17}$$

总的耗尽区厚度为

$$W = x_n + x_P \tag{2.18}$$

在耗尽区对公式(2.15)积分，得到内建电势为

$$V_{bi} = -\int_{-x_P}^{x_n}\mathscr{E}(x)dx = -\int_{-x_P}^{o}\mathscr{E}(x)dx\,|_{P侧} - \int_{O}^{x_n}\mathscr{E}(x)dx\,|_{n侧}$$

$$= \frac{qN_Ax_P^2}{2\varepsilon_s} + \frac{qN_Dx_n^2}{2\varepsilon_s} = \frac{1}{2}\mathscr{E}_mW \tag{2.19}$$

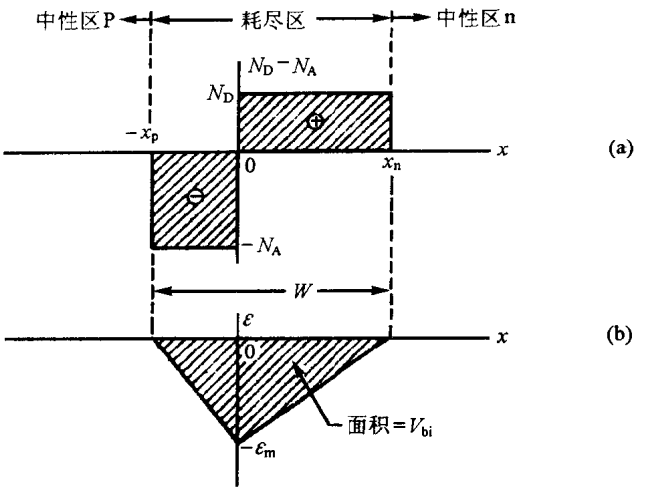


图 2.3 (a) 热平衡下耗尽区的空间电荷分布;(b) 电场分布,阴影区面积等于内建电势

联立(2.16)至(2.19)式,得到

$$x_P = \left[\frac{2\varepsilon_s}{q}\cdot\frac{N_D}{(N_A+N_D)N_A}V_{bi}\right]^{\frac{1}{2}} \tag{2.20}$$

$$x_n = \left[\frac{2\varepsilon_s}{q}\cdot\frac{N_A}{(N_A+N_D)N_D}V_{bi}\right]^{\frac{1}{2}} \tag{2.21}$$

$$W = \left[\frac{2\varepsilon_s}{q}\cdot\frac{N_A+N_D}{N_AN_D}V_{bi}\right]^{\frac{1}{2}} \tag{2.22}$$

实际问题中经常遇到 p 区和 n 区掺杂浓度相差悬殊的情况,即所谓单边突变结的情况,例如 p^+-n 结中 $N_A \geqslant N_D$,这时有

$$x_P \simeq 0 \tag{2.23}$$

$$W \simeq x_n = \left(\frac{2\varepsilon_s}{qN_D}V_{bi}\right)^{\frac{1}{2}} \tag{2.24}$$

$$\mathscr{E}_m = \frac{qN_DW}{\varepsilon_s} \tag{2.25}$$

$$\psi(x) = |\mathscr{E}_m|\left(x-\frac{x^2}{2W}\right) \tag{2.26}$$

公式(2.26)已经规定了 p 型中性区为电势零点,即 $\psi(0)=0$。

2. 线性缓变结

热平衡时的空间电荷分布和电场分布如图 2.4 所示,泊松方程为

$$\frac{d^2\psi}{dx^2} = -\frac{d\mathscr{E}}{dx} = -\frac{q}{\varepsilon_s}ax \tag{2.27}$$

a 是杂质浓度梯度(单位为 cm^{-4}),电场的边界条件为

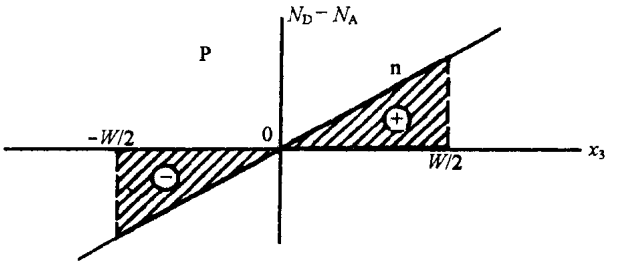
$$\mathscr{E}|_{x=\pm\frac{W}{2}} = 0 \tag{2.28}$$

从以上二式得到电场分布

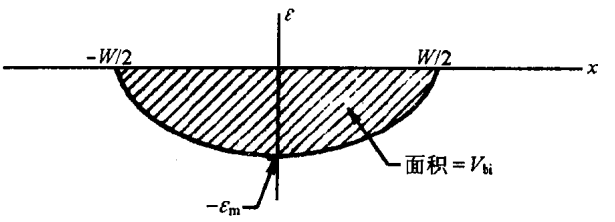
$$\mathscr{E}(x) = -\frac{qa}{\epsilon_s}\left[\frac{\left(\frac{W}{2}\right)^2 - x^2}{2}\right] \tag{2.29}$$

最大电场强度在 $x=0$ 处，大小为

$$\mathscr{E}_m = \frac{qa}{8\epsilon_s}W^2 \tag{2.30}$$



(a) 空间电荷分布



(b) 电场分布

图 2.4　热平衡时的线性缓变结

W 为耗尽区宽度，内建电势为

$$V_{bi} = -\int_{-\frac{W}{2}}^{\frac{W}{2}}\mathscr{E}(x)\mathrm{d}x = \frac{qa}{12\epsilon_s}W^3 \tag{2.31}$$

由上式得到线性缓变结的耗尽区宽度为

$$W = \left[\frac{12\epsilon_s}{qa}V_{bi}\right]^{\frac{1}{3}} \tag{2.32}$$

另一方面，如果把线性缓变结看成由无数薄层组成，每一薄层的掺杂浓度均匀，则类似于突变结，内建电势可由耗尽区边缘$\left(-\frac{W}{2}\text{及}\frac{W}{2}\right)$处的杂质浓度表示为

$$\begin{aligned} V_{bi} &= \frac{k_B T}{q}\ln\left[\frac{\left(a\frac{W}{2}\right)\left(a\frac{W}{2}\right)}{n_i^2}\right] \\ &= \frac{2k_B T}{q}\ln\left(\frac{aW}{2n_i}\right) \end{aligned} \tag{2.33}$$

将公式(2.32)和(2.33)联立，消去 W，就得到线性缓变结的内建电势 V_{bi}，它是杂质浓度梯度 a 的函数。对于 Si 和 GaAs 线性缓变结，内建电势和杂质浓度梯度的关系示于图 2.5。

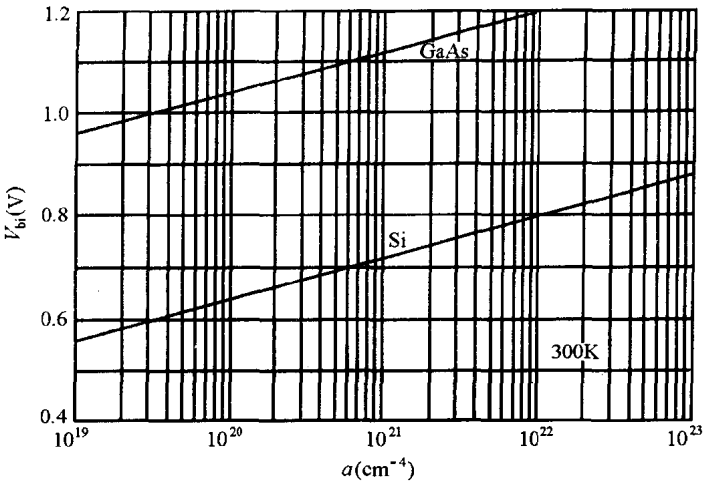


图 2.5　Si 和 GaAs 线性缓变结的内建电势与杂质梯度的关系

2.2.2　非平衡情形

有外加电压时，p-n 结不再处于平衡状态，能带图如图 2.6 所示。由于一般情形下势垒区内的载流子浓度很低，电阻很大，而势垒区外的p 区或 n区载流子浓度较高，电阻较小，外加电压基本上都降落在势垒区上。这样，如图 2.6(b)所示，对 p 区加一个相对于 n 区为正的电压(正向偏置) V_F 时，n 区的电位相对 p 区降低，势垒高度由热平衡时的 V_{bi}降低为 $V_{bi}-V_F$；反

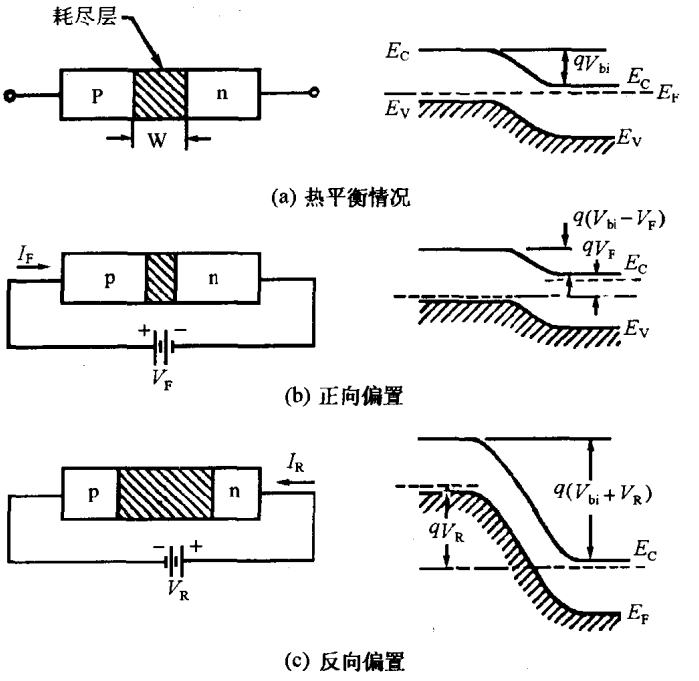


图 2.6　在不同偏置条件下 p-n 结的耗尽层宽度和能带图的示意图

之，如图 2.6(c)所示，对 n 区加一个相对于 p 区为正的电压(反向偏置) V_R 时，势垒高度增加到 $V_{bi}+V_R$。所以在一定条件下，把与势垒高度相关的物理量(例如耗尽层宽度)的表示式，从热平衡推广到非平衡情形，具体做法是把 V_{bi} 换成 $V_{bi}-V$，正向偏置时 V 是正的，反向偏置时 V 是负的。

2.2.3 耗尽层电容(势垒电容)

p-n 结的耗尽区宽度随外加电压改变，因此它所包含的电离杂质数目随外加电压改变，即耗尽区内的正、负电荷量随外加电压改变。这种电压变化引起电荷量变化的电容效应称为耗尽层电容，或势垒电容。耗尽区是一个由电离施主和电离受主组成的正负电荷量相等的偶极层，设 Q 为单位面积耗尽区内正的(或负的)电荷量，则单位面积耗尽层电容定义为

$$C_j = \frac{dQ}{dV} \tag{2.34}$$

对于任意掺杂分布的 p-n 结，在外加电压增量 dV 的作用下，耗尽区宽度变化产生的增量电荷 dQ 将是在耗尽区边缘。由电中性条件，耗尽区 n 型和 p 型两侧边缘附近都有增量电荷 dQ，但符号相反，所以耗尽区内的电场强度增量为

$$d\mathscr{E} = dQ/\varepsilon_s \tag{2.35}$$

所以耗尽区的电压增量为

$$dv \simeq W \cdot dQ/\varepsilon_s \tag{2.36}$$

根据式(2.34)给出的定义，由式(2.36)得到，对任意掺杂分布的 p-n 结，单位面积耗尽层电容为

$$C_j = \frac{\varepsilon_s}{W} \tag{2.37}$$

ε_s 是半导体介电常数，W 是耗尽区宽度，由于 dQ/dV 表示微分电容，上式只是小信号情形下成立。

由式(2.37)可以直接得到突变结和线性缓变结的耗尽层电容，分别为

$$C_j = \left[\frac{q\varepsilon_s N_A N_D}{(N_A+N_D)}\right]^{\frac{1}{2}}(V_{bi}-V)^{-\frac{1}{2}} \tag{2.38}$$

和

$$C_j = \left[\frac{qa\varepsilon_s^3}{12}\right]^{\frac{1}{3}}(V_{bi}-V)^{-\frac{1}{3}} \tag{2.39}$$

对于单边突变结，式(2.38)简化为

$$C_j = \sqrt{\frac{q\varepsilon_s N_B}{2(V_{bi}-V)}} \tag{2.40}$$

N_B 是轻掺杂一例的杂质浓度。

在应用(2.36)式写出(2.38)至(2.40)式时，耗尽区宽度 W 由热平衡时的表示式通过以 $V_{bi}-V$ 代替 V_{bi} 得到，这在反向偏压或小的正向偏压下是合理的；但对于大的正向偏压，势垒区内载流子浓度和电离杂质浓度可以相比拟，耗尽近似不再成立，上述推广得到的结果也就不再有效。

2.3 直流特性

2.3.1 理想伏-安特性

下面根据下述条件推导 p-n 结的电流-电压方程：(1) 全部外加电压降落在耗尽区上，耗尽区外半导体为电中性；(2) 小注入；(3) 耗尽区内不存在复合或产生电流；(4) 半导体非简并。

非简并条件下，电子和空穴都服从玻耳兹曼分布，即

$$n = n_i e^{(E_{Fn}-E_i)/k_B T} \tag{2.41}$$

$$p = n_i e^{(E_i-E_{Fp})/k_B T} \tag{2.42}$$

E_{Fn}、E_{Fp}分别为电子、空穴的准费米能级。非平衡时，半导体中的导带电子和价带空穴应看作两个不同的体系，各自有自己的费米能级。电子和空穴的浓度乘积为

$$pn = n_i^2 e^{(E_{Fn}-E_{Fp})/k_B T} \tag{2.43}$$

如果耗尽区宽度显著小于载流子的扩散长度，则可认为 E_{Fn}和 E_{Fp}平直通过耗尽区，这时在耗尽区内(包括耗尽区边界)有 $E_{Fn}-E_{Fp}=qV$，V 是外加电压，所以

$$pn = n_i^2 e^{qV/k_B T} \tag{2.44}$$

以脚标 n 和 p 表示半导体的类型，以脚标"0"标志热平衡，则小注入条件下的多子浓度分别为 $n_n \simeq n_{n0}$和 $p_p \simeq p_{p0}$，在耗尽区边界的少子浓度分别为

$$n_p(-x_p) = \frac{n_i^2}{p_p} e^{qV/k_B T} \simeq n_{p0} e^{qV/k_B T} \tag{2.45}$$

$$p_n(x_n) \simeq p_{n0} e^{qV/k_B T} \tag{2.46}$$

在 n 型中性区($x \geqslant x_n$)内空穴分布的连续方程为

$$\frac{\partial p_n}{\partial t} = D_p \frac{\partial^2 p_n}{\partial x^2} - \frac{p_n - p_{n0}}{\tau_p} \tag{2.47}$$

D_p 和 τ_p 分别为空穴的扩散系数和寿命，上式右边第一项来自空穴流动产生积累的速率，第二项来自空穴被复合的速率，由于中性区内的电场被忽略，空穴只有扩散运动，稳态时$\frac{\partial p_n}{\partial t}=0$，故有

$$\frac{d^2 p_n}{dx^2} - \frac{p_n - p_{n0}}{D_p \tau_p} = 0 \tag{2.48}$$

对突变结，$dp_{n0}/dx=0$，上式可改写为

$$\frac{d^2(p_n - p_{n0})}{dx^2} - \frac{p_n - p_{n0}}{L_p^2} = 0 \tag{2.49}$$

$L_p=(D_p\tau_p)^{\frac{1}{2}}$为空穴的扩散长度，n 型中性区中空穴分布的边界条件为 $x=W_n$，W_n 为 n 区长度)时 $p_n-p_{n0}=0$，$x=x_n$ 时 $p_n-p_{n0}=p_{n0}\left[\exp\left(\frac{qV}{k_B T}\right)-1\right]$，方程(2.49)满足上述边界条件的解为

$$p_n - p_{n0} = p_{n0}(e^{qV/k_B T} - 1) \frac{\sinh\left(\frac{W_n - x}{L_p}\right)}{\sinh\left(\frac{W_n - x_n}{L_p}\right)} \tag{2.50}$$

对 n 区很长($W_n \geqslant L_p$)的 p-n 结,上式简化为

$$p_n - p_{n0} = p_{n0}(e^{qV/k_BT} - 1)e^{-(x-x_n)}/L_p \tag{2.51}$$

空穴注入引起的电流密度为

$$J_p = -qD_p \frac{d(p_n - p_{n0})}{dx}\bigg|_{x_n} = \frac{qD_p}{L_p} p_{n0}(e^{qV/k_BT} - 1) \tag{2.52}$$

同样,在 p 型中性区,有

$$n_p - n_{p0} = n_{p0}(e^{qV/k_BT} - 1)e^{(x+x_p)/L_n} \tag{2.53}$$

$L_n = \sqrt{D_n \tau_n}$是电子的扩散长度,电子注入 p 型中性区引起的电流密度为

$$J_n = qD_n \frac{q(n_p - n_{p0})}{dx}\bigg|_{-x_p} = \frac{qD_n}{L_n} n_{p0}(e^{qV/k_BT} - 1) \tag{2.54}$$

图 2.7 表示注入的少数载流子分布和电子、空穴扩散电流。注入的少数载流子离开边界后,便不断和多数载流子复合,在 n 区中的空穴扩散电流和在 p 区中的电子扩散电流都随距离指数衰减,其特征长度分别为 L_p 和 L_n。

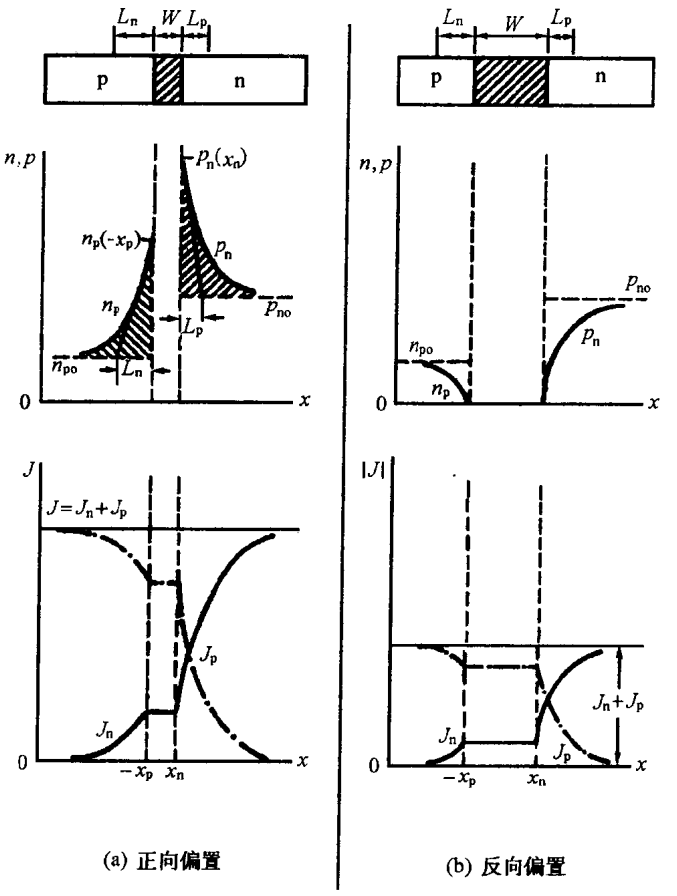


图 2.7 注入的少数载流子分布和电子、空穴扩散电流

在忽略耗尽区内的产生或复合电流的条件下,由式(2.52)和(2.54)相加,就得到通过 p-n

结的总电流：

$$J = J_s(e^{qV/k_BT} - 1) \tag{2.55}$$

$$J_s = \frac{qD_p p_{n0}}{L_p} + \frac{qD_n n_{p0}}{L_n} \tag{2.56}$$

方程(2.55)称为理想二极管方程，J_s 是饱和电流密度。表示式(2.56)适合于长的 p-n 结。对于中性区长度小于少数载流子扩散长度的短 p-n 结，式中的 L_p 和 L_n 应当分别用 $W_n - x_n$ 和 $W_p - x_p$ 代替。

2.3.2 产生-复合效应

理想二极管方程只能与实际情形定性符合，偏压很小时发生偏离的主要原因是耗尽区内的产生－复合过程。根据半导体物理，通过间接复合中心复合率为

$$U = \frac{pn - n_i^2}{\tau_p(n + n_t) + \tau_n(p + p_t)} \tag{2.57}$$

式中，$pn = n_i^2 e^{qV/k_BT}$，$n_t = n_i e^{(E_i - E_t)/k_BT}$；$E_t$ 是复合中心的能级位置；τ_p 和 τ_n 是载流子寿命：

$$\tau_p = \frac{1}{\sigma_n v_{th} N_t} \tag{2.58}$$

$$\tau_n = \frac{1}{\sigma_n v_{th} N_t} \tag{2.59}$$

σ_n 和 σ_p 是复合中心对电子和空穴的俘获截面，v_{th} 是电子和空穴的热运动速度，N_t 是复合中心浓度，根据上述，对于 $\sigma_n = \sigma_p = \sigma_0$ 的简单情形，复合率 U 可表述为

$$U = \frac{\sigma_0 v_{th} N_t n_i^2 (e^{qV/k_BT} - 1)}{n + p + 2n_i \cosh \dfrac{(E_i - E_t)}{k_BT}} \tag{2.60}$$

反向偏置时，$pn < n_i^2$，电子和空穴的产生过程占支配地位，如果 $p < n_i$ 和 $n < n_i$，则根据式(2.66)，产生率为

$$G = -U = \frac{\sigma_0 v_{th} N_t n_i}{2\cosh \dfrac{(E_i - E_t)}{k_BT}} = \frac{n_i}{\tau_g} \tag{2.61}$$

τ_g 是载流子浓度恢复到平衡态的弛豫时间，通常称为载流子寿命。在这里代表产生寿命，由式(2.61)可以看出，产生率在 $E_t = E_i$ 时最大，而当偏离 E_i 时指数下降，即只有能级 E_t 靠近禁带中央的中心才对产生有显著贡献。实际上，无论产生或复合，最有效的是能级在禁带中央附近的那些中心。

耗尽区内载流子产生而形成的电流密度为

$$J_g = \int_0^W qG\,dx \simeq \frac{qn_i W}{\tau_g} \tag{2.63}$$

J_g 和 n_i 成正比，J_s 和 n_i^2 成正比，所以对于 n_i 大的半导体，例如 Ge，室温下扩散电流占支配地位，反向电流遵从理想二极管方程；对 n_i 小的半导体，例如 Si 和 GaAs，室温下耗尽区产生电流可能占支配地位。

正向偏置时，$pn > n_i^2$，耗尽区内起支配作用的是电子和空穴的复合过程。对于 $E_t = E_i$ 的

情形，式(2.60)简化为

$$U = \frac{\sigma_0 v_{th} N_t n_i^2 (e^{qV/k_B T} - 1)}{n + p + 2n_i} \tag{2.64}$$

可以证明，耗尽区中 $n = p$(等于 $n_i e^{qV/2k_B T}$)的地方复合率有最大值：

$$U_{max} = \sigma_0 v_{th} N_t \frac{n_i^2 (e^{qV/k_B T} - 1)}{2n_i (e^{qV/2k_B T} + 1)} \tag{2.65}$$

当 $V > \frac{3k_B T}{q}$ 时有

$$U_{max} = \frac{n_i}{2\tau_r} e^{qV/2k_B T} \tag{2.66}$$

其中 $\tau_r = 1/\sigma_0 v_{th} N_t$ 表示有效复合寿命。

耗尽区载流子复合引起的电流密度为

$$J_r = \int_0^W qU \mathrm{d}x \simeq \frac{qn_i W}{2\tau_r} e^{qV/2k_B T} \tag{2.67}$$

对于 p^+-n 结，总的正向电流密度为

$$J_F = q \sqrt{\frac{D_p}{\tau_p}} \frac{n_i^2}{N_D} e^{qV/k_B T} + \frac{qn_i W}{2\tau_r} e^{qV/2k_B T} \tag{2.68}$$

通常，实验结果可用下述经验公式表示：

$$J_F \sim e^{qV/\eta k_B T} \tag{2.69}$$

式中的 η 称为理想因子，其数值在 1 和 2 之间，小电流时，复合电流占支配地位，$\eta = 2$；较大电流时，理想扩散电流占支配地位，$\eta = 1$。

2.3.3 大注入效应

所谓大注入，是正向偏压下注入的少数流子浓度可以与多数载流子浓度相比，即在 n 区有 $p_n(x_n) \simeq n_n$，p 区有 $n_p(-x_p) \simeq p_p$，将这一条件代入 $pn = n_i^2 e^{qV/k_B T}$，就得到大注入时的边界条件

$$p_n(x_n) = n_i e^{qV/2k_B T} \tag{2.70}$$

和

$$n_p(-x_p) = n_i e^{qV/2k_B T} \tag{2.71}$$

另方面，在 n 型或 p 型中性区的少数载流子扩散区域内，非平衡载流子浓度随离开耗尽区边界的距离衰减，稳定时，这种不均匀分布靠扩散区域内的自建电场来维持。大注入时，注入扩散区域内的少数载流子数量很大，这种自建电场作用不能忽略，从多数载流子电流为零的条件可以求得这一自建电场，以 n 型中性区为例，在空穴扩散的区域内，自建电场 $\mathscr{E}$ 满足 $J_n = 0$ 的条件，即

$$q\mu_n n_n \mathscr{E} + qD_n \frac{\mathrm{d}n_n}{\mathrm{d}x} = 0 \tag{2.72}$$

故有

$$\mathscr{E} = -\frac{k_B T}{q} \frac{1}{n_n} \frac{\mathrm{d}n_n}{\mathrm{d}x} \tag{2.73}$$

注入的空穴电流密度为

$$J_p(x_n) = \left[q\mu_p p_n \mathscr{E} - qD_p \frac{dp_n}{dx}\right]_{x_n} = -qD_p\left[\frac{p_n}{n_n}\frac{dn_n}{dx} + \frac{dp_n}{dx}\right]_{x_n}$$

$$= -q(2D_p)\left.\frac{dp_n}{dx}\right|_{x_n} \tag{2.74}$$

上式表明,在大注入条件下,自建电场对少数载流子(空穴)漂移的作用相当于空穴的扩散系数增加一倍,这一结论同样适合于注入 p 区的电子电流。

综合上述,大注入电流的表示式可以在小注入分析结果的基础上得到,只需以 $2D_p$ 和 $2D_n$ 分别代替 D_p 和 D_n,以 $n_i\exp(qV/2k_BT)$代替 $p_{n0}\exp(qV/k_BT)$或 $n_{p0}\exp(qV/k_BT)$,所得电流大体上正比于 $\exp(qV/2k_BT)$,故大电流下电流随电压增加的速度减小。

大电流时,在串联电阻(包括中性区的电阻和非理想欧姆接触的电阻)上的电压降落也不能忽略,以 R 表示串联电阻,I 表示流过 p-n 结的电流,则加在势垒上的电压实际只有 $V-IR$,其中 V 是电源电压,理想扩散电流的表示式应为

$$I = I_s e^{q(V-IR)/k_BT} \tag{2.75}$$

它下降一个因子 e^{IR/k_BT},这在电流很大时十分显著。

2.3.4 温度效应

无论是正向偏置,还是反向偏置,扩散电流和产生一复合电流都强烈地依赖于温度,根据前面的讨论,扩散电流为

$$I_{扩散} = I_s(e^{qV/k_BT} - 1) \tag{2.76}$$

式中,I_s 为饱和电流,

$$I_s = qA\left(\frac{D_n n_{p0}}{L_n} + \frac{D_p p_{n0}}{L_p}\right) = qA\left(\frac{D_n}{L_n N_A} + \frac{D_p}{L_p N_D}\right) n_i^2 \tag{2.77}$$

A 为 p-n 结面积。复合和产生电流分别为

$$I_{复合} = qA\frac{Wn_i}{2\tau_r}e^{qV/2k_BT} \tag{2.78}$$

$$I_{产生} = qA\frac{Wn_i}{\tau_g} \tag{2.79}$$

反向偏置情形 扩散电流与产生电流之比为

$$\frac{I_{扩散}}{I_{产生}} \simeq \frac{\tau_g}{W}\left(\frac{D_n}{L_n N_A} + \frac{D_p}{L_p N_D}\right) n_i \tag{2.80}$$

随温度增加,n_i 增大,$I_{扩散}$将占优势。下面根据式(2.76)来讨论温度对反向电流的影响。当反向偏压 $V > 3k_BT/q$ 时,有

$$I_{扩散} = I_s \sim n_i^2 \sim T^3 e^{-E_{g0}/k_BT} \tag{2.81}$$

E_{g0}为绝对零度时的禁带宽度,对于 Si, $E_{g0} \simeq 1.17$eV。由式(2.81)可得

$$\frac{1}{I_s}\frac{dI_s}{dT} \simeq \frac{E_{g0}}{k_BT^2} \tag{2.82}$$

k_B 约等于 8.62×10^{-5}eV/K。可见,在室温(300K)附近,对于硅 p-n 结,温度每增加 1℃,I_s 相

对增加 15%，即温度每增加 6℃，就会使反向电流增加一倍。

正向偏置情形 若 $V>3k_BT/q$，则有

$$\frac{I_{扩散}}{I_{复合}}=\frac{2\tau_r}{W}\left(\frac{D_n}{L_nN_A}+\frac{D_p}{L_pN_D}\right)n_ie^{qV/2k_BT}\sim e^{-(E_{g0}-qV)/2k_BT} \tag{2.83}$$

上式说明，偏压增加或温度增加时，$I_{扩散}$对$I_{复合}$的比率增加，由图 2.8 可以看出，高于室温时，不大的正向电压(~0.3V)就使 $I_{扩散}$占优势。所以，下面我们仅根据 $I_{扩散}$来讨论温度对正向伏安特性的影响。

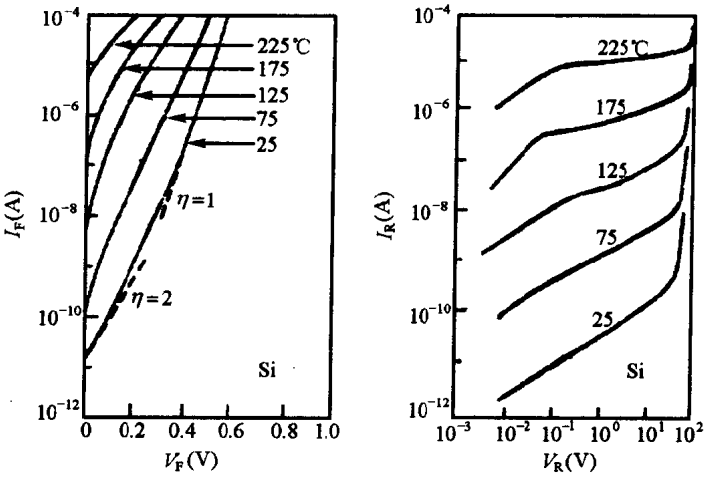


图 2.8 硅二极管的伏-安特性和温度的关系

根据 $I_{扩散}\simeq I_s\exp(qV/k_BT)$，若维持 I 不变，则结电压 V 随温度的变化为

$$\left.\frac{dV}{dT}\right|_I=\frac{V}{T}-\frac{k_BT}{q}\left(\frac{1}{I_s}\frac{dI_s}{dT}\right)=\frac{V-E_{g0}/q}{T} \tag{2.84}$$

若维持结电压不变，则正向电流 I 随温度的变化率为

$$\frac{1}{I}\left(\frac{dI}{dT}\right)_V=\frac{1}{I_s}\cdot\frac{dI_s}{dT}-\frac{qV}{k_BT^2}=\frac{E_{g0}-qV}{k_BT^2} \tag{2.85}$$

对于硅 p-n 结，典型的工作电压为 $V=0.6$ 伏，在室温附近，每增加 10℃，电流约增加一倍；电压的变化率约 -2mV/℃，所以结电压随温度的变化十分灵敏，这一特性被用来精确测温和控温。

2.4 交流小信号特性；扩散电容

设 p-n 结正向偏置，偏压为 V_0，在结上又叠加一个频率 ω 的交流小电压，$V_1e^{j\omega t}$($V_1\ll V_0$)，则结上总电压为

$$V(t)=V_0+V_1e^{j\omega t} \tag{2.86}$$

仍从 n 型中性区开始分析。在边界 $x=x_n$ 处的空穴浓度为

$$p_n(x_n,t)=p_{n0}e^{qV/k_BT}\simeq p_{n0}e^{qV_0/k_BT}+\frac{qV_1}{k_BT}p_{n0}e^{qV_0/k_BT}e^{j\omega t}$$

$$= p_{ns}(x_n) + p_{n1}(x_n)e^{j\omega t} \tag{2.87}$$

上式的第二步用到了 $V_1 \ll V_0$ 的小信号条件，式中，$p_{ns}(x_n)$代表 $p_n(x_n, t)$的稳定值，$p_{n1}(x_n)$则是 $p_n(x_n, t)$随时间变化部分的振幅。从这一结果可以看出，在小信号条件下，n 型中性区内的空穴分布可以表示成

$$p_n(x, t) = p_n(x) + p_{n1}(x)e^{j\omega t} \tag{2.88}$$

上式右边第一项代表稳定分布，第二项代表与时间有关的分布，$p_n(x, t)$满足连续方程

$$\frac{\partial p_n(x, t)}{\partial t} = D_p \frac{\partial^2 p_n(x, t)}{\partial x^2} - \frac{p_n(x, t) - p_{n0}}{\tau_p} \tag{2.89}$$

将式(2.88)代入上式，并注意 $D_p \frac{\partial^2 p_{ns}}{\partial x^2} - \frac{p_{ns} - p_{n0}}{\tau_p} = 0$，便可得到一个关于交流振幅 $p_{n1}(x)$的微分方程：

$$\frac{d^2 p_{n1}(x)}{dx^2} - \frac{p_{n1}(x)}{L_p'^2} = 0 \tag{2.90}$$

式中

$$L'_p = L_p / \sqrt{1 + j\omega\tau_p} \tag{2.91}$$

方程(2.90)所应满足的边界条件为：$x = W_n$ 时 $p_{n1} = 0$；$x = x_n$ 时 $p_{n1}(x_n) = \frac{qV_1}{k_B T} p_{n0} e^{qV_0/k_B T}$。满足边界条件的解为

$$p_{n1}(x) = p_{n1}(x_n) \frac{\sinh\left(\frac{W_n - x}{L'_p}\right)}{\sinh\left(\frac{W_n - x_n}{L'_p}\right)} \tag{2.92}$$

W_n 是 n 区的长度。对于长 p-n 结，n 型中性区内空穴分布的交流分量为

$$p_{n1}(x, t) = p_{n1}(x_n)e^{-(x - x_n)/L'_p} e^{j\omega t} \tag{2.93}$$

注入 n 区的空穴电流中的交流分量为

$$J_{p1}(t) = -qD_p \frac{\partial p_n(x, t)}{\partial x}\bigg|_{x_n} = \frac{qD_p p_{n1}(x_n)}{L'_p} e^{j\omega t} \tag{2.94}$$

同理，注入 p 区的电子电流中的交流分量为

$$J_{n1}(t) = \frac{qD_n n_{p1}(-x_p)}{L'_n} e^{j\omega t} \tag{2.95}$$

式中

$$L'_n = L_n / \sqrt{1 + j\omega\tau_n} \tag{2.96}$$

$$n_{p1}(-x_p) = \frac{qV_1}{k_B T} n_{p0} e^{qV_0/k_B T} \tag{2.97}$$

将式(2.94)和(2.95)相加，就得到通过 p-n 结的交流电流密度为

$$\begin{aligned} J_1(t) &= \left[\frac{qD_p p_{n1}(x_n)}{L'_p} + \frac{qD_n n_{p1}(-x_p)}{L'_n}\right] e^{j\omega t} \\ &= \frac{qV_1}{k_B T}\left[\frac{qD_p p_{n0}}{L_p / \sqrt{1 + j\omega\tau_p}} + \frac{qD_n n_{p0}}{L_n / \sqrt{1 + j\omega\tau_n}}\right] e^{qV_0/k_B T} \cdot e^{j\omega t} \end{aligned} \tag{2.98}$$

交流导纳 Y(等于 $J_1(t)/V_1e^{j\omega t}$)可表示为

$$Y = G_d + j\omega C_d \tag{2.99}$$

上式右方第一项 G_d 表示电导;第二项是电容性的,C_d 为扩散电容,它来源于 n 型和 p 型中性区的电容效应,中性区内存储有等量的过剩电子电荷和空穴电荷,它们的数量受结电压控制,当频率很低($\omega\tau_p, \omega\tau_n \leqslant 1$)时,将 $J_1(t)$ 的振幅作泰勒展开,只保留到一次项,即得扩散电导为

$$G_d = \frac{q}{k_B T}\left[\frac{qD_p p_{n0}}{L_p} + \frac{qD_n n_{p0}}{L_n}\right]e^{qV_0/k_B T} \tag{2.100}$$

扩散电容为

$$C_d = \frac{q}{k_B T}\left[\frac{qL_p p_{n0}}{2} + \frac{qL_n n_{p0}}{2}\right]e^{qV_0/k_B T} \tag{2.101}$$

C_d 与外加信号的频率 ω 有关,器件物理中常以乘积 $\omega\tau$ 来划分频率的高低,其中 τ 是非平衡载流子的寿命,亦即存储电荷再分布的弛豫时间。$\omega\tau_p, \omega\tau_n \ll 1$ 的条件标志外加信号变化周期远大于存储电荷再分布时间的低频情形,C_d 由式(2.101)表示;而对 $\omega\tau_p, \omega\tau_n \gg 1$ 的高频情形,存储电荷跟不上结电压的变化,C_d 很小(C_d 随 $\sqrt{\omega\tau}$ 下降)。另方面,C_d 随直流偏压增加,所以,在低频和正向偏压下,扩散电容特别重要。

在许多应用中,总是根据在使用条件下半导体器件各部分的物理作用,用电阻、电容、电流源和电压源等组成一定的电路来等效器件的功能。根据前面的讨论,对理想 p-n 结,小信号等效电路如图 2.9 所示。图中,C_j 是耗尽层电容,C_d 是扩散电容,G 是电流通过器件的电导,C_j,C_d 和 G 分别由式(2.37),(2.101)和(2.100)乘以器件面积 A 得到。

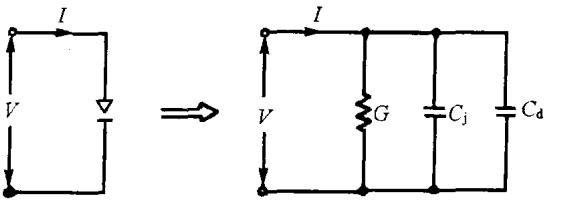


图 2.9 p-n 结小信号等效电路

2.5 电荷存储和反向恢复时间

正向偏置时,电子从 n 区注入到 p 区,空穴从 p 区注入到 n 区。少数载流子一旦注入,便吸引多数载流子并与之复合,最终形成非平衡载流子浓度随距离衰减的稳定分布。这些非平衡载流子导致 p-n 结内的电荷积累,即等量的过剩电子电荷和过剩空穴电荷的存储。当结上外加偏压突然反向时,这些存储电荷不能立即去除,即需要经过一定时间 p-n 结才能达到反偏状态,这个时间称为反向恢复时间,其间存储电荷和电流等随时间的变化说明 p-n 结的反向瞬态特性。

2.5.1 存储电荷

为简单起见,考虑一个长的 p^+-n 结,并将 $x=0$ 取作边界。在 n 型中性区内对过剩空穴积分便得到存储电荷量,有

$$Q = qA\int_0^\infty (p_n - p_{n0})dx = qA\int_0^\infty p_{n0}(e^{qV_F/k_B T} - 1)e^{-x/L_p}dx$$

$$= qAL_p p_{n0}(e^{qV_F/k_B T} - 1) \tag{2.102}$$

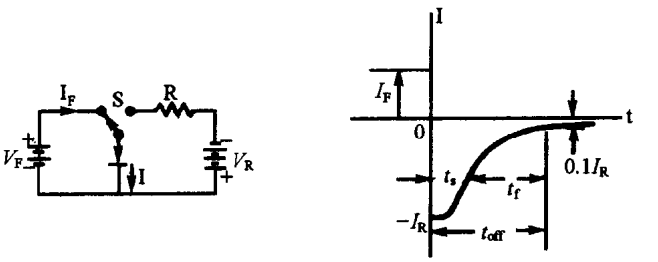
A 是结面积，V_F 是外加正向偏压，利用 $I_F = \frac{qAD_p}{L_p} p_{n0}(e^{qV_F/k_BT} - 1)$ 及 $L_p = \sqrt{D_p\tau_p}$，上式可改写为

$$Q = \tau_p I_F \tag{2.103}$$

即存储电荷量是注入电流和少数载流子寿命的乘积。注入电流越大，进入中性区的少数载流子越多；少数载流子寿命越长，它们在复合之前扩散进中性区越远，存储的电荷也就越多。

2.5.2 瞬态特性和反向恢复时间

图 2.10 表示 p-n 结的瞬态特性，其中图 2.10(a)表示一个能够改变 p-n 结偏置方向的简单电路。正向时，有一电流 I_F 流过 p-n 结。在 $t=0$ 时刻，开关 S 突然掷到右边，此时通过 p-n 结的电流并不是立即变为反向饱和电流 I_s，而是先经过一个较大的恒定反向电流 I_R 的阶段，然后再逐渐衰减到 I_s，如图 2.10(b)所示。即 p-n 结的反向瞬变过程可以分为电流恒定和电流衰减两个阶段，相应的瞬变时间分别以 t_s 和 t_f 表示。t_s 称为存储时间；t_f 称为下降时间，定义为从 I_R 开始衰减到 $0.1I_R$ 所经过的时间。$t_{off} = t_s + t_f$ 即为反向恢复时间，它比偏压从反向突变为正向的瞬变时间长得多。



(a) 基本开关电路　　(b) 由正向偏置到反向偏置的瞬态响应

图 2.10　p-n 结瞬态特性

下面采用电荷控制模型，对长 p^+-n 结的情形，近似计算瞬变时间。反偏刚开始的一段时间($0 < t < t_s$)内，p^+-n 结处于反向低阻状态，电流基本恒定，由外电路决定：

$$I_R \simeq V_R/R \tag{2.104}$$

R 为偏置电阻，存储电荷通过外电路流走(反向抽取)和自身复合减少，即

$$-\frac{dQ}{dt} = I_R + Q/\tau_p \tag{2.105}$$

上述方程的解为

$$Q(t) = \tau_p(I_F + I_R)e^{-t/\tau_p} - \tau_p I_R \tag{2.106}$$

得到上式时已用了 $t=0$ 时 $Q = \tau_p I_F$ 的初始条件。令 $t = t_s$ 时存储电荷完全消失，即 $Q(t_s) = 0$，由此得到存储时间为

$$t_s = \tau_p \ln[1 + I_F/I_R] \tag{2.107}$$

经过时间 t_s 以后，可认为 p-n 结已经进入反向高阻状态。

严格分析瞬变过程应根据空穴分布的连续方程(2.47)，对于 p^+-n 结，结果为

$$\operatorname{erf}\sqrt{\frac{t_s}{\tau_p}} = (1 + I_R/I_F)^{-1} \tag{2.108}$$

$$\operatorname{erf}\sqrt{\frac{t_s}{\tau_p}}+\frac{\exp(-t_f/\tau_p)}{\sqrt{\pi(t_f/\tau_p)}}=1+0.1\left(\frac{I_R}{I_F}\right) \tag{2.109}$$

其中 erfx 为误差函数，表示为

$$\operatorname{erf}x=\frac{2}{\sqrt{\pi}}\int_0^x e^{-u^2}du \tag{2.110}$$

当比值 I_R/I_F 很小时，对于 n 型材料长度 W_n 远大于扩散长度 L_p（$W_n\geqslant L_p$）的长 p^+-n 结，反向恢复时间可近似表示为

$$t_s+t_f\simeq\frac{\tau_p}{2}\left(\frac{I_R}{I_F}\right)^{-1} \tag{2.111}$$

对于 $W_n\leqslant L_p$ 的情形，则近似为

$$t_s+t_f\simeq\frac{W_n^2}{2D_p}\left(\frac{I_R}{I_F}\right)^{-1} \tag{2.112}$$

由上述分析可知，对于高速开关，要求 τ_p 很小。如果引进能级靠禁带中央的复合中心（例如硅中掺金），则可大大降低少数载流子寿命 τ_p。此外，金属-半导体二极管（肖特基势垒二极管）具有高速开关特性，这是因为它是多数载流子承担电流输运的器件，电荷存储效应可以忽略。

2.6 结的击穿

加到 p-n 结上的反向电压超过某一临界值 V_B 时，p-n 结发生击穿，产生很大的电流。V_B 称为击穿电压。击穿本身不是破坏性的，但必须通过外电路对最大电流加以限制，以避免 p-n 结过热烧毁。p-n 结击穿有两种重要机制：隧道效应和雪崩倍增，如图 2.11 所示。碰撞电离引起的雪崩倍增使大多数半导体器件的工作电压受到限制；另方面，雪崩倍增现象也可以利用（已被用在产生微波功率和探测光信号等方面）。所以本节将详细讨论雪崩击穿，而对隧道击穿将在第六章讨论。

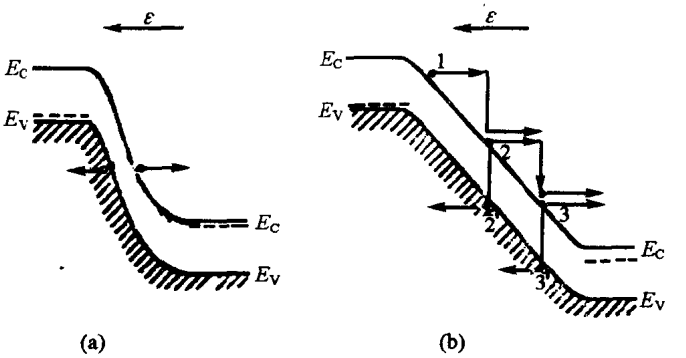


图 2.11 (a)隧道效应；(b)雪崩倍增

2.6.1 隧道击穿

在 p 区和 n 区掺杂浓度相当高（$>5\times10^{17}cm^{-3}$）的 p-n 结中，当外加反向电场强度达到

10^6v/cm 或更高时，由于空间电荷区宽度很小，p 区价带占据态上的电子有很大几率穿过禁带进入 n 区导带的空态，如图 2.11(a)所示。这种因为电子隧道穿透效应在强电场下迅速增加而产生大电流的现象，称为隧道击穿。对硅和砷化镓 p-n 结，击穿电压小于 $4E_g/q$（E_g 为禁带宽度）时，通常是隧道击穿；击穿电压大于 $6E_g/q$ 时，则是雪崩击穿。击穿电压在 $4E_g/q$ 和 $6E_g/q$ 之间时，是这两种击穿的混合。

2.6.2 雪崩击穿

对于掺杂浓度不太高的 p-n 结，例如 p^+-n 结，当反向偏压很大时，热产生电子能够在短距离内从强电场中获得足够能量同晶格原子碰撞，破坏晶格键，产生电子-空穴对，这一过程称为碰撞电离。新产生的电子和空穴再从电场获得能量，进一步产生电子-空穴对。这种连锁过程称为雪崩倍增。由雪崩倍增而产生大电流的现象称为雪崩击穿。

1. 雪崩击穿条件

图 2.12 表示 p-n 结耗尽区内的雪崩倍增。初始电子电流 I_{n0}从耗尽区左边（$x=0$）流入，若耗尽区内电场强度大到引起雪崩倍增，则电子电流 $I_n(x)$随距离增大，流出耗尽区时增大到 $I_n(W)=M_nI_{n0}$，M_n 称为电子的倍增因子，

$$M_n=\frac{I_n(W)}{I_{n0}} \tag{2.113}$$

初始空穴电流 I_{p0}则从耗尽区右边（$x=W$）流入，流出耗尽区时增大到 $I_p(0)=M_pI_{p0}$。空穴倍增因子为

$$M_p=\frac{I_p(0)}{I_{p0}} \tag{2.114}$$

稳态时通过 p-n 结的总电流 $I=I_n+I_p$ 为常量。

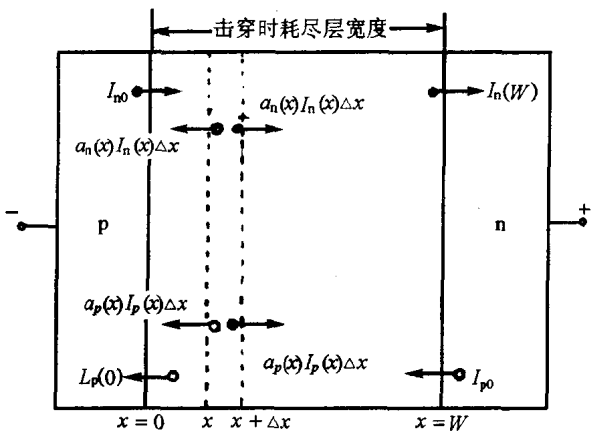


图 2.12 p-n 结耗尽区内的雪崩倍增

在 x 处增加的电子流等于 dx 范围内电子和空穴的倍增作用之和，即

$$d\left(\frac{I_n}{q}\right)=\alpha_n\frac{I_n}{q}dx+\alpha_p\frac{I_p}{q}dx \tag{2.115}$$

或

$$\frac{dI_n}{dx}+(\alpha_p-\alpha_n)I_n=\alpha_pI \tag{2.116}$$

式中 α_n 和 α_p 分别是电子和空穴的电离率，即电子或空穴在单位距离内通过碰撞电离产生的电子-空穴对数目。α_n 和 α_p 都强烈依赖于电场强度，所以一般情形下是位置 x 的函数。同样，对空穴流的增加可以写出

$$\frac{dI_p}{dx}+(\alpha_p-\alpha_n)I_p=-\alpha_nI \tag{2.117}$$

方程(2.116)和(2.117)是 $dy/dx+p(x)y=Q(x)$类型的微分方程，可以分别利用 $I_n(W)=$

$M_n I_{n0}$和 $I_p(0) = M_p I_{p0}$的边界条件求解①。下面只讨论 $\alpha_n = \alpha_p = \alpha$ 的简单情形。这时，两个方程分别简化为 $dI_n/dx = \alpha I$ 和 $dI_p/dx = -\alpha I$，从而有

$$I_n(x) = I\int_0^x \alpha dx + I_{n0} \tag{2.118}$$

和

$$I_p(x) = -I\int_x^W \alpha dx + I_{p0} \tag{2.119}$$

以上二式相加，并注意 $I_n(x) + I_p(x) = I$，得到

$$1 - \frac{1}{M} = \int_0^W \alpha dx \tag{2.120}$$

式中 M 是倍增因子，定义为

$$M = I/I_0 \tag{2.121}$$

它表示雪崩倍增以后的总电流 I 与初始电流 $I_0(= I_{n0} + I_{p0})$之比，当外加反向偏压使 $\int_0^W \alpha dx \to 1$ 时，$M \to \infty$，I 增加到无限大。这样，我们得到雪崩击穿条件：

$$\int_0^W \alpha dx = 1 \tag{2.122}$$

2. 雪崩击穿电压

由实验测得的电离率对电场强度的依赖关系，利用雪崩击穿条件，可以计算出发生雪崩击穿的临界电场强度 $\mathscr{E}_c$(击穿时的最大电场强度)。由于电离率和电场强度之间有比较复杂的函数关系，而且除某些特殊情形(例如 p-i-n 结)外，电场强度是位置 x 的函数，计算十分复杂。为了应用方便，已经把计算结果画成图表，例如图 2.13；或者表示成经验公式，例如：

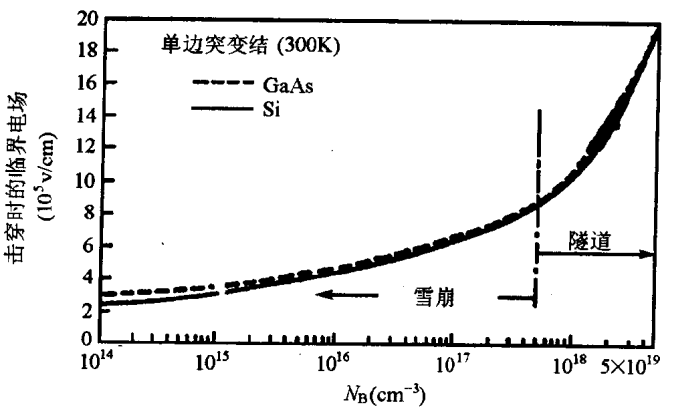


图 2.13 Si, GaAs 单边突变结击穿临界电场和衬底掺杂浓度的关系

① 可以证明

$$1 - \frac{1}{M_n} = \int_0^W \alpha_n \left[\exp\left(-\int_0^x (\alpha_n - \alpha_p)dx'\right)\right]dx \quad 和 \quad 1 - \frac{1}{M_p} = \int_0^W \alpha_p \left[\exp\left(-\int_x^W (\alpha_p - \alpha_n)dx'\right)\right]dx$$

$$\mathscr{E}_c \simeq 1.1 \times 10^7 \left(\frac{q}{\varepsilon_s}\right)^{\frac{1}{2}} \left(\frac{E_g}{1.1}\right)^{\frac{3}{4}} N_B^{\frac{1}{8}} \quad \text{（单边突变结）} \tag{2.123}$$

和

$$\mathscr{E}_c \simeq 1.5 \times 10^8 \left(\frac{q}{\varepsilon_s}\right)^{\frac{1}{3}} \left(\frac{E_g}{1.1}\right)^{\frac{4}{5}} a^{1/15} \quad \text{（线性缓变结）} \tag{2.124}$$

其中，E_g 是半导体材料的禁带宽度，N_B 是轻掺杂一侧的杂质浓度，a 是杂质浓度梯度。

根据临界电场，利用前述关于耗尽区电压、最大电场强度同耗尽区宽度的关系，可以计算击穿电压 V_B 和击穿时耗尽区宽度 X_{mB}。对单边突变结，有：

$$V_B = \frac{\varepsilon_s \mathscr{E}_c^2}{2q} \cdot \frac{1}{N_B} \tag{2.125}$$

$$X_{mB} = \left[\frac{2\varepsilon_s V_B}{qN_B}\right]^{\frac{1}{2}} \tag{2.126}$$

对线性突变结，有：

$$V_B = \frac{4}{3} \mathscr{E}_c^{\frac{3}{2}} \left[\frac{2\varepsilon_s}{q}\right]^{\frac{1}{2}} a^{-\frac{1}{2}} \tag{2.127}$$

$$X_{mB} = \left[\frac{12\varepsilon_s V_B}{qa}\right]^{\frac{1}{3}} \tag{2.128}$$

对于一定的半导体材料，把 E_g、ε_s 和 q 的数据代入，可直接表示出 p-n 结的击穿电压同掺杂浓度或杂质浓度梯度的函数关系，例如图 2.14 所示的情形。图 2.15 表示扩散 p-n 结的击穿电压与杂质浓度梯度 a 及原材料掺杂浓度 N_B 的关系。对于大的 a 和低的 N_B，击穿电压主要由衬底掺杂浓度决定，相当于单边突变结，如图 2.15 的底线所示；对于小的 a 和高的 N_B，击穿电压主要由杂质浓度梯度决定，相当于线性缓变结，如图 2.15 的平行线所示。

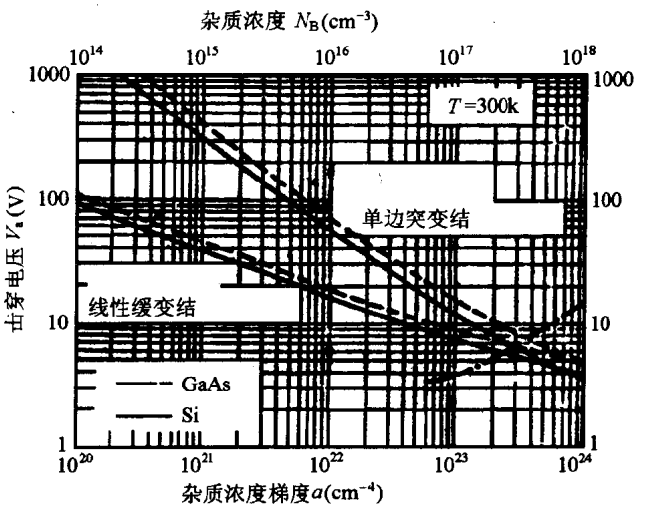


图 2.14　Si, GaAs 雪崩击穿电压和单边突变结及线性缓变结的杂质浓度的关系（点划线表示开始出现隧道击穿机制）

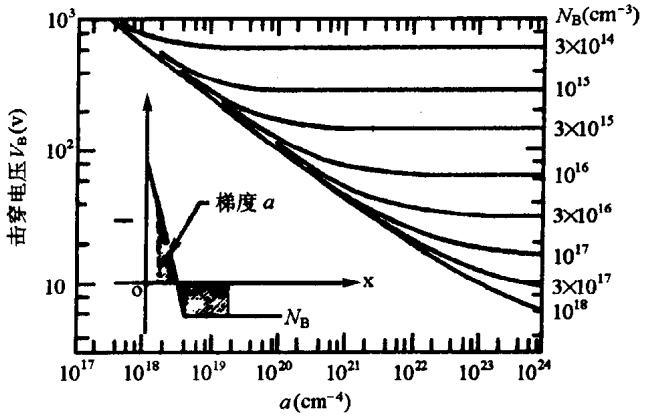


图 2.15　300 K 时，扩散 p-n 结的击穿电压与原材料杂质浓度 N_B、杂质浓度梯度 a 的关系

实际 p-n 结的击穿电压受许多因素的影响，首先要考虑半导体厚度 W。当 $W < X_{mB}$（击穿时耗尽区宽度）时，器件将穿通，也就是击穿前耗尽区已扩散到了衬底，如图 2.16 中的插图所示。随着反向电压进一步增加，器件将发生击穿。由于对一定的半导体材料和 p-n 结形状，临界击穿电场 $\mathscr{E}_c$ 基本上是常数，所以穿通 p-n 结的击穿电压 V'_B 满足

$$\frac{V'_B}{V_B} = \frac{\text{图中阴影面积}}{\frac{1}{2}\mathscr{E}_c X_{mB}} = 1 - \frac{(X_{mB} - W)^2}{X_{mB}^2} \tag{2.129}$$

这一结果说明低掺杂半导体层厚度 W 比 W_{mB}小得越多，击穿电压 V'_B 比由电阻率决定的雪崩击穿电压 V_B 低得越多，图 2.16 给出 p^+-n-n^+ 结或 p^+-p-n^+ 结的击穿电压与半导体层厚度的关系曲线，图中 W 代表 n 区或 p 区的厚度，实线为 $W > X_{mB}$时的曲线，也就是图 2.14 给出

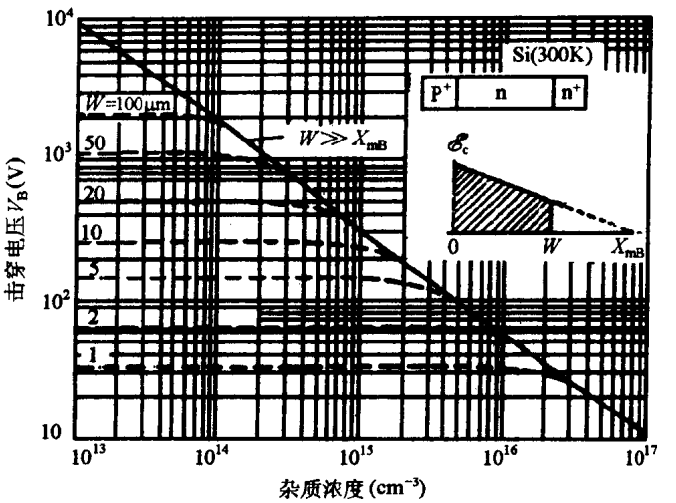


图 2.16　300K 时硅 p^+-n-n^+ 结或 p^+-p-n^+ 结击穿电压与低掺杂区杂质浓度中厚度的关系

的单边突变结击穿电压与杂质浓度的关系曲线，虚线为 $W < X_{mB}$ 时的击穿电压与杂质浓度的关系曲线。从图可以看出，同样杂质浓度，当 W 比 X_{mB} 小得多时，击穿电压基本上由 W 的数值决定。

击穿电压还和结面的形状有关。通过半导体表面绝缘层的窗口(通常是矩形窗口)扩散形成 p-n 结时，杂质既向下作纵向扩散，又向侧面作横向扩散，如图 2.17 所示。因此，结的侧面是圆柱形，底部是平面结，尖角处是球面，如图 2.18 所示。曲率半径小的地方电场集中而首先击穿，所以球面区或柱面区决定器件的击穿电压，它将低于前面按照平面结模型得出的数值。

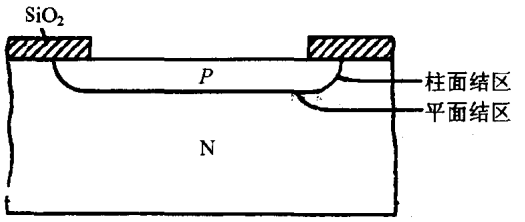


图 2.17　扩散掩膜的边缘结形成弯曲

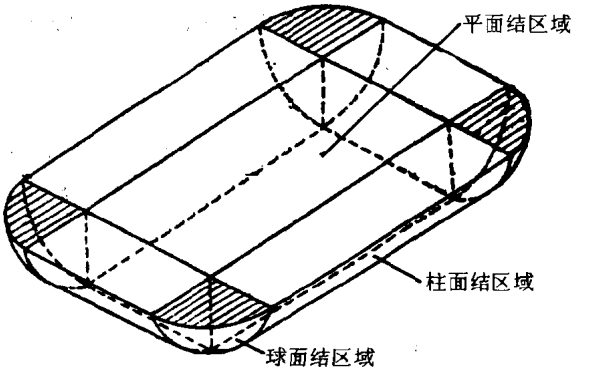


图 2.18　通过矩形掩膜扩散成柱面区和球面区

对硅的柱面或球面单边突变结，击穿电压可表示为

$$V_B \simeq 6 \times 10^{13} N_B^{-3/4} \{[(n+1+\gamma)\gamma^n]^{\frac{1}{n+1}} - \gamma\} \qquad (1.130)$$

式中，$\gamma = X_j / X_{mB}$，X_j 是扩散结深，X_{mB} 是平面 p-n 结击穿时的耗尽区宽度；$n=1$ 用于柱面结，$n=2$ 用于球面结。计算结果如图 2.19 所示，实线代表平面结的情况。可以看出，当结的半径 X_j 减小时击穿电压明显减小，特别是对掺杂浓度低的球面结更是如此。为了消除结面弯曲

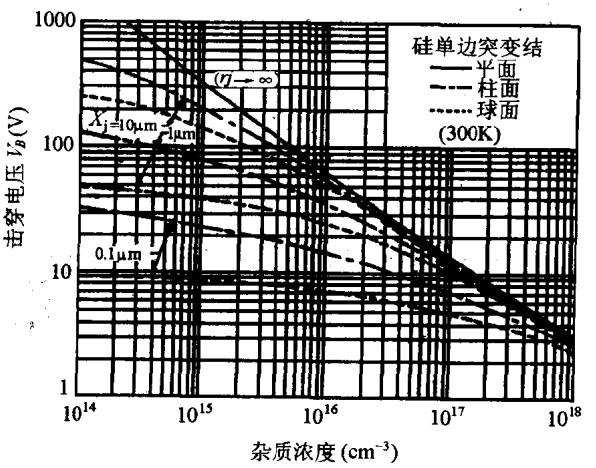


图 2.19　不同结深情况下，硅平面结、柱面结、球面结的击穿电压与原材料杂质浓度的关系

造成局部电场集中的影响，工艺上可采用高浓度扩散环或台面结来切断横向扩散，也可以用深扩散结的方法，还可以采用磨角(正斜角)的方法。所谓正斜角是指斜面横截面积由高浓度侧向低浓度侧逐渐减小。为维持耗尽区内正、负空间电荷数量相等，高掺杂一侧的空间电荷区宽度略微减小，而低掺杂一侧的空间电荷区宽度明显增大，结果表面耗尽区宽度大于体内耗尽区宽度，而导致表面电场减弱。

此外，V_B 随温度增加(正温度系数)，这是因为温度高时晶格振动强烈，载流子同晶格碰撞损失的能量增加，使得从电场积累能量的速度减慢，只有在更强的电场下才能具有碰撞电离和雪崩倍增所需的能量。

第三章　双极型晶体管

双极型晶体管是最先(1947 年)出现的三端半导体器件,具有放大作用,高速性能尤其突出。近二十年来,金属-氧化物-半导体场效应晶体管(MOSFET)技术迅速发展,双极型晶体管的突出地位受到了严重挑战,但它在诸如高速计算机、火箭和卫星、现代通信和电力系统方面仍是关键性器件。而且,双极技术目前也有了突破性进展,双极型晶体管有希望保持其在电路工作速度方面的领先地位。

关于双极型晶体管的理论、工艺和应用,已经有大量著作,本章只是叙述双极型晶体管的基本原理和一般特性,这些知识对理解其应用和学习双极集成电路都是必不可少的基础。

3.1　基 本 原 理

3.1.1　基本结构

双极型晶体管由二个 p-n 结组成,即一薄层 p 型半导体夹在二层 n 型半导体中间,或者反过来,n 型半导体夹在 p 型半导体中间。前者称为 n-p-n 晶体管,后者称为 p-n-p 晶体管。这两种晶体管的基本结构及其在电路中的表示符号示于图 3.1 中,图中箭头表示正常工作条件(放大状态)下的电流方向。

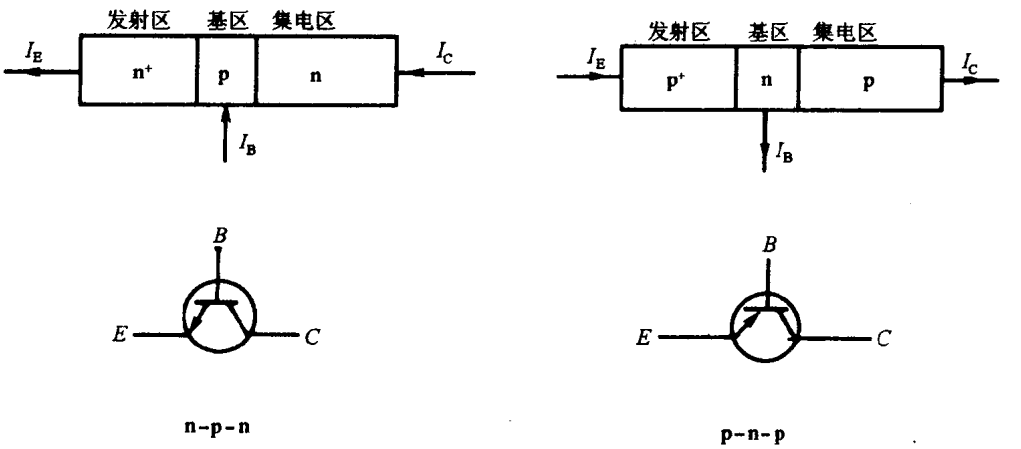


图 3.1　晶体管结构的简化模型和代表符号

本章用 n-p-n 结构来叙述双极型晶体的机理。但不论采用何种结构,结果是可以互换的,只要改变电压极性和导电类型,就可以换到另一种结构出发得到的结果。

3.1.2　放大工作状态

图 3.2 表示 n-p-n 晶体管偏置在放大状态时的情形。这时,基区-发射区结(发射结)必须

加正向偏压，基区-集电区结(集电结)必须加大的反向偏压。

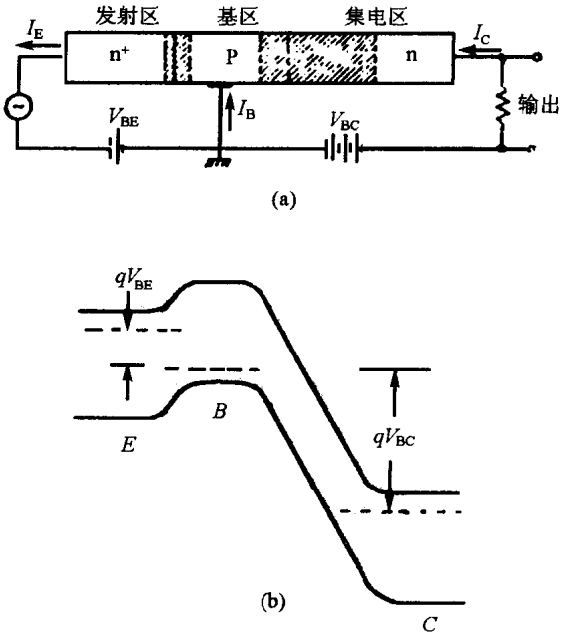


图 3.2　n-p-n 晶体管共基极放大电路(a)和能带图(b)

由于发射结正偏，其上的电势降低 qV_{BE}，电子将从发射区向基区注入，空穴将从基区向发射区注入。所以基区有过剩电子，发射区有过剩空穴，过剩电子或空穴的浓度不仅取决于发射结偏压的大小，还取决于发射区或基区的掺杂浓度。集电结强反偏，本身只有一个很小的反向电流，但当基区宽度十分小(远远小于电子扩散长度)时，从发射区注入的电子除少数被复合外，其余大多数能到达集电结耗尽区边缘，然后被扫入集电区，集电极电流因此基本上等于发射极电流中的电子电流。如果在集电极回路中接入较大的负载电阻，就可以将信号放大。

如果基区宽度远大于电子(少子)的扩散长度，晶体管仅仅是两个背靠背的 p-n 结，不可能有放大作用。

3.1.3　电流增益

图 3.3 表示工作在放大状态时 n-p-n 晶体管内的电流。I_{En}是从发射区注入到基区中的电子电流，I_{Cn}是被集电区收集到的电子电流，I_{En}和 I_{Cn}是晶体管内主要的电流分量。I_{Ep}是从基区注入到发射区的空穴电流，通常比 I_{En}小得多。I_{ER}是发射结耗尽区内的复合电流。I_{BR}是基区内的复合而必须补充的空穴电流。I_{CBO}是流过集电结的反向电流，主要是集电结附近热产生的空穴被扫入基区形成的电流。

由图 3.3，晶体管的终端电流，即发射极电流 I_E，集电极电流 I_C 和基极电流 I_B 如下：

$$I_E = I_{En} + I_{Ep} + I_{ER} \tag{3.1}$$

$$I_C = I_{Cn} + I_{CBO} \tag{3.2}$$

$$I_B = I_{Ep} + I_{ER} + (I_{En} - I_{Cn}) - I_{CBO} \tag{3.3}$$

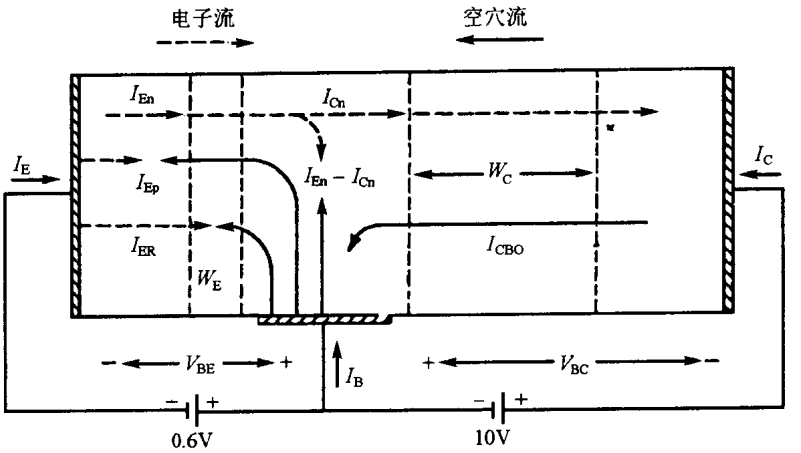


图 3.3　放大工作时 n-p-n 晶体管内的电流

若规定流入晶体管的电流方向为正，则有

$$I_E + I_C + I_B = 0 \tag{3.4}$$

双极型晶体管的重要参数之一是共基极直流短路电流增益，定义为

$$\alpha_0 \equiv \frac{I_{Cn}}{I_E} \tag{3.5}$$

将式(3.1)代入，有

$$\alpha_0 \equiv \frac{I_{Cn}}{I_{En}+I_{Ep}+I_{ER}} = \left[\frac{I_{En}}{I_{En}+I_{Ep}+I_{ER}}\right]\left[\frac{I_{Cn}}{I_{En}}\right] = \gamma\alpha_T \tag{3.6}$$

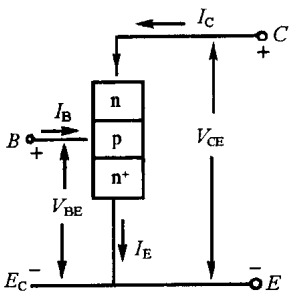


图 3.4　n-p-n 晶体管的共发射极接法

γ 称为发射效率，

$$\gamma \equiv \frac{I_{En}}{I_{En}+I_{Ep}+I_{ER}} \tag{3.7}$$

α_T 称为基区传输因子，

$$\alpha_T \equiv \frac{I_{Cn}}{I_{En}} \tag{3.8}$$

利用 α_0，集电极电流可表示为

$$I_C = \alpha_0 I_E + I_{CBO} \tag{3.9}$$

在许多电路应用中，常利用共发射极接法(见图 3.4)，为了表示相应的电流增益，将式(3.9)改写为

$$I_C = \alpha_0(I_C + I_B) + I_{CBO} \tag{3.10}$$

或

$$I_C = \frac{\alpha_0}{1-\alpha_0} I_B + \frac{I_{CBO}}{1-\alpha_0} \tag{3.11}$$

定义发射极直流短路电流增益

$$\beta_0 \equiv \frac{\Delta I_C}{\Delta I_B} = \frac{\alpha_0}{1-\alpha_0} \tag{3.12}$$

这里,增量 ΔI_C 和 ΔI_B 的含义只不过是 I_C 和 I_B 中不考虑 I_{CBO}的贡献(因为 I_{CBO}不受 V_{BE}控制),而并非小量。

令

$$I_{CEO} = \frac{I_{CBO}}{1 - \alpha_0} \tag{3.13}$$

I_{CEO}是基极开路($I_B=0$)的集电极-发射极漏电流。需要注意,基极开路并非发射结上没有电压,实际上,由于强反偏的集电结对发射区的电子有一定吸引作用,发射结势垒降低,即 $V_{BE} \gtrsim 0$。所以,发射结有微弱的注入作用,使 $I_{CEO} \gg I_{CBO}$。利用 β_0 和 I_{CEO},集电极电流可表示为

$$I_C = \beta_0 I_B + I_{CEO} \tag{3.14}$$

3.2 双极型晶体管的直流特性

3.2.1 理想晶体管的电流

假定发射区、基区和集电压都是均匀掺杂,杂质分布及耗尽区如图 3.5 所示,图中画斜线的区域为耗尽区。并假定小注入,耗尽区内没有产生-复合电流,器件中不存在串联电阻。

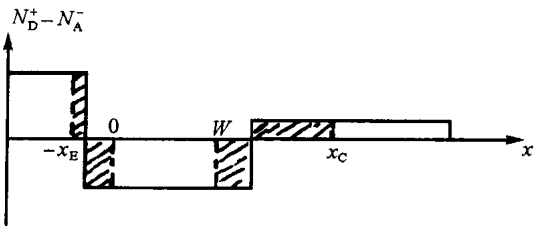


图 3.5 各区均匀掺杂 n-p-n 晶体管的杂质分布

在中性基区($0 \leqslant x \leqslant W$)内,电子分布满足没有电场的稳态连续方程:

$$D_n \frac{d^2 n_p}{dx^2} - \frac{n_p - n_{p0}}{\tau_n} = 0 \tag{3.15}$$

及下述边界条件:

$$n_p(O) = n_{p0} e^{qV_{BE}/k_B T} \tag{3.16}$$

$$n_p(W) = n_{p0} e^{qV_{BC}/k_B T} \tag{3.17}$$

解之,得到基区内的电子(少数载流子)分布为

$$n_p(x) = n_{p0} + n_{p0}(e^{qV_{BE}/k_B T} - 1)\left[\frac{\sinh\left(\frac{W-x}{L_n}\right)}{\sinh\left(\frac{W}{L_n}\right)}\right] + n_{p0}(e^{qV_{BC}/k_B T} - 1)\left[\frac{\sinh(x/L_n)}{\sinh(W/L_n)}\right] \tag{3.18}$$

令 $W/L_n \ll 1$,流过发射结的电子电流为

$$I_{En} = A\left[qD_n \frac{dn_p(x)}{dx}\bigg|_{x=0}\right]$$

$$\simeq -qA\frac{D_n n_{p0}}{W}\left[(e^{qV_{BE}/k_BT}-1)-\frac{1}{\cosh(W/L_n)}(e^{qV_{BC}/k_BT}-1)\right] \tag{3.19}$$

负号表示实际的电流方向与所选坐标方向相反。同理,流过集电结的电流为

$$I_{Cn}=A\left[qD_n\frac{dn_p(x)}{dx}\Big|_{x=W}\right]$$

$$\simeq -qA\frac{D_n n_{p0}}{W}\left[\frac{1}{\cosh(W/L_n)}(e^{qV_{BE}/k_BT}-1)-(e^{qV_{BC}/k_BT}-1)\right] \tag{3.20}$$

发射区和集电区内的少子(空穴)分布可用类似在基区内采用的方法得到,这时的边界条件为

$$p_E(-x_E)=p_{E0}e^{qV_{BE}/k_BT} \tag{3.21}$$

和

$$p_C(x_C)=p_{C0}e^{qV_{BC}/k_BT} \tag{3.22}$$

假定发射区宽度 W_E 和集电区宽度 W_c 远大于相应的扩散长度 L_{pE}和 L_{pC},则有

$$p_E(x)=p_{E0}+p_{E0}(e^{qV_{BE}/k_BT}-1)e^{(x+x_E)/L_{pE}},\ x\leqslant -x_E \tag{3.23}$$

和

$$p_C(x)=p_{C0}+p_{C0}(e^{qV_{BC}/k_BT}-1)e^{-(x-x_C)/L_{pc}},\ x\geqslant x_C \tag{3.24}$$

注入发射区和集电区和空穴电流分别为

$$I_{Ep}=-qA\frac{D_{pE}p_{E0}}{L_{pE}}(e^{qV_{BE}/k_BT}-1) \tag{3.25}$$

和

$$I_{cp}=-qA\frac{D_{pc}p_{C0}}{L_{pc}}(e^{qV_{BC}/k_BT}-1) \tag{3.26}$$

3.2.2 电流基本方程

一个双极型晶体管可以有四种工作状态,取决于基区-发射区结和基区-集电结所加直流偏压的极性。除前面已提到的放大状态外,还有三种可能的工作状态:饱和、截止和反向放大。在饱和工作状态下,两个结都处于正向偏置,晶体管处于小偏置电压、大输出电流情况,基区电子(少子)分布的边界条件为 $n_p(0)=n_{p0}e^{qV_{BE}/k_BT}$和 $n_p(W)=n_{p0}e^{qV_{BC}/k_BT}$,其中 V_{BE}和 V_{BC}均大于零。在截止工作状态下,两个结都处于反向偏置,基区电子分布的边界条件变为 $n_p(0)=n_p(W)=0$,晶体管处于大偏置电压、小输出电流情况。在反向工作状态下,基区-发射区结处于反向偏置,基区-集电区结处正向偏置,也就是说将器件反向使用,这时的发射效率通常低于放大状态。如图 3.6 所示。

由式(3.19),(3.25),(3.20)和(3.26)得出对 n-p-n 晶体管所有工作状态普遍适合的电流基本方程(规定流入晶体管的电流方向为正):

$$I_E=I_{En}+I_{Ep}=a_{11}(e^{qV_{BE}/k_BT}-1)+a_{12}(e^{qV_{BC}/k_BT}-1) \tag{3.27}$$

和

$$I_C=I_{Cn}+I_{cp}=a_{21}(e^{qV_{BE}/k_BT}-1)+a_{22}(e^{qV_{BC}/k_BT}-1) \tag{3.28}$$

式中,

$$a_{11} = -qA\left(\frac{D_n}{N_B W} + \frac{D_{pE}}{N_E L_E}\right) n_i^2 \tag{3.29}$$

$$a_{12} \simeq qA\frac{D_n}{N_B W} n_i^2 \tag{3.30}$$

$$a_{21} = a_{12} \tag{3.31}$$

$$a_{22} = -qA\left(\frac{D_n}{N_B W} + \frac{D_{pc}}{N_c L_c}\right) n_i^2 \tag{3.32}$$

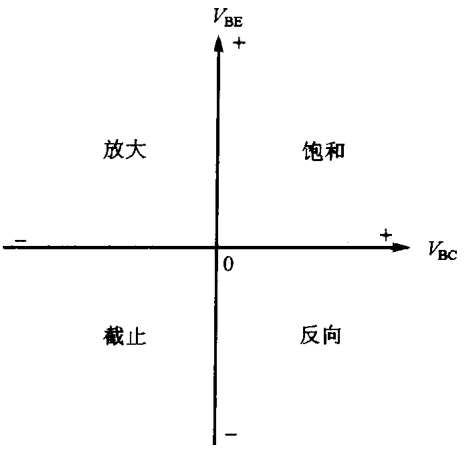


图 3.6　n-p-n 晶体管的工作状态和结偏压极性的关系

3.2.3　放大状态

n-p-n 晶体管工作在放大状态的偏置条件是

$$V_{BE} > 0,\ V_{BC} < 0 \tag{3.33}$$

下面把这一条件应用到前述有关的表达式中，写出晶体内的少子分布和电流，并讨论电流增益。

基区内的电子分布(设 $W/L_n \ll 1$)为

$$n_p(x) \simeq n_{p0} e^{qV_{BE}/k_B T}(1 - x/W) = n_p(0)\left(1 - \frac{x}{W}\right) \tag{3.34}$$

分布曲线近似为一直线，发射区和集电区内的空穴分布分别为

$$p_E(x) = p_{E0} + p_{E0}(e^{qV_{BE}/k_B T} - 1)e^{(x+x_E)/L_{pE}},\ x \leqslant -x_E \tag{3.35}$$

和

$$p_C(x) \simeq p_{C0} - p_{C0} e^{-(x-x_c)/L_{pc}},\ x \geqslant x_c \tag{3.36}$$

图 3.7 表示工作在放大状态时 n-p-n 晶体管内的少子分布，这里着重要讨论的是基区内的少子存储电荷 Q_B:

$$Q_B = -qA\int_0^W [n_p(x) - n_{p0}]dx \simeq \frac{-qAWn_p(0)}{2} \tag{3.37}$$

放大状态下晶体管内电流的表达式将分别为

$$I_{En} \simeq -qA\frac{D_n n_{p0}}{W}e^{qV_{BE}/k_BT} \tag{3.38}$$

$$I_{Cn} \simeq -qA\frac{D_n n_{p0}}{W}\cdot\frac{1}{\cosh(W/L_n)}e^{qV_{BE}/k_BT} \tag{3.39}$$

$$I_{Ep} \simeq -qA\frac{D_{pE} p_{E0}}{L_E}e^{qV_{BE}/k_BT} \tag{3.40}$$

和

$$I_{cp} \simeq -qA\frac{D_{pc} p_{c0}}{L_c} \tag{3.41}$$

集电极电流则为

$$I_c \simeq -qA\frac{D_n n_{p0}}{W}e^{qV_{BE}/k_BT} \tag{3.42}$$

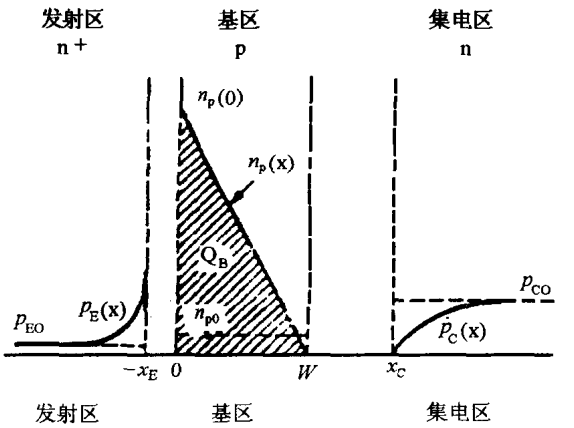


图 3.7　放大状态时 n-p-n 晶体管内的少子分布

当发射结电压 V_{BE}不是十分小时，其耗尽区内的复合电流 I_{ER}可忽略，于是发射效率为

$$\gamma = \frac{I_{En}}{I_{En}+I_{Ep}} = \frac{1}{1+\dfrac{D_{pE}}{D_n}\dfrac{W}{L_E}\dfrac{p_{E0}}{n_{p0}}} \tag{3.43}$$

或

$$\gamma = \frac{1}{1+\dfrac{D_{pE}}{D_n}\dfrac{W}{L_E}\dfrac{N_B}{N_E}} \tag{3.44}$$

为了提高 γ，应尽可能降低掺杂比 N_B/N_E，主要是尽可能提高 N_E，这就是要采用重掺杂 n^+ 发射区的理由，但 N_E 也不能太大，否则会引起半导体禁带宽度变窄和俄歇复合显著，二者都使注入发射区的空穴电流增加，γ 反而下降。同时，根据其他方面的考虑，N_B 也不能太小。所以用减小 N_B/N_E 的方法来提高 γ 是十分有限的。要进一步提高 γ，必须采用新型的发射结结构，其中最重要的是本章最后两节介绍的异质结双极晶体管和多晶硅发射极晶体管。

基区传输因子为

$$\alpha_T = \frac{I_{Cn}}{I_{En}} \simeq \frac{1}{\cosh\left(\frac{W}{L_n}\right)} \simeq 1 - \frac{W^2}{2L_n^2} \tag{3.45}$$

3.2.4 非理想现象分析

前面关于理想晶体管的分析能近似描述实际器件的电流-电压特性，但也有一定偏差，下面讨论引起偏差的主要因素或物理效应。

1. 缓变基区

理想晶体管假定发射区、基区和集电区的杂质分布都是均匀的。然而在实际制造晶体管时，通常采用所谓外延平面工艺，用热扩散或离子注入的方法向外延衬底掺杂，基区的杂质分布并不均匀，而是存在很大的浓度梯度。如图 3.8 所示。基区内杂质的浓度梯度使空穴向集电区扩散；而在热平衡条件下，中性基区将有一个内建电场来抵消扩散电流，也就是说电场把空穴推向发射区，使之不产生净电流。但是这一内建电场将增强电子的运动。在放大的偏置条件下，所注入的电子不仅有扩散运动，还有由基区内建电场所引起的漂移运动。内建电场的主要作用是缩短注入电子渡越基区所需要的时间，从而改善晶体管的高频特性；另一与此有关的作用是改善基区的传输因子，因为减小渡越基区所需的时间就可能减少电子在基区的复合。

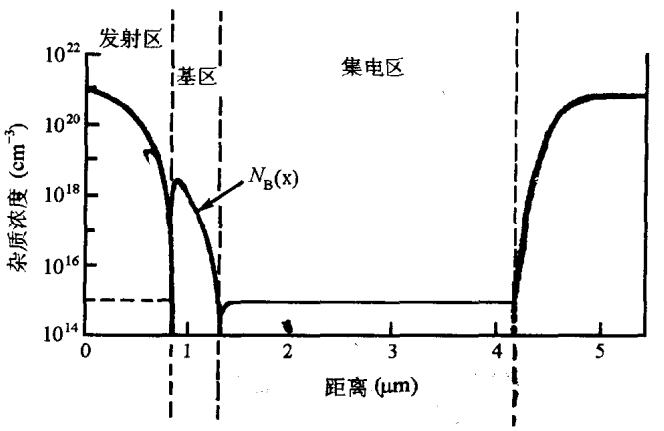


图 3.8 扩散型双极晶体管的杂质分布

以 $\mathscr{E}$ 表示基区内建电场的强度，$\mathscr{E}$ 可以通过基区内空穴(多数载流子)电流为零的条件

$$qp_p\mu_p\mathscr{E} - qD_p\frac{dp_p}{dx} = 0$$

得到，于是有

$$\mathscr{E} = \frac{k_BT}{q}\frac{1}{p_p(x)}\cdot\frac{dp_p(x)}{dx} \tag{3.46}$$

将式(3.46)代入基区内电子电流的表示式

$$qn_p(x)\mu_n\mathscr{E} + qD_n\frac{dn_p(x)}{dx} = J_n$$

并利用爱因斯坦关系 $\mu_n=(q/k_BT)D_n$,得出

$$\frac{dn_p(x)}{dx}+\frac{n_p(x)}{p_p(x)}\frac{dp_p(x)}{dx}=\frac{I_n}{qAD_n} \tag{3.47}$$

A 为基区的横截面积。如果忽略基区复合(即 I_n 与 x 无关并等于 I_{En}),并设 D_n 为常数,以及在放大状态下 $n_p(W)=0$,可得

$$n_p(x)=\frac{I_{En}}{qAD_np_p(x)}\int_W^x p_p(x)dx \tag{3.48}$$

和

$$I_{En}=-\frac{qAD_nn_i^2e^{qV_{BE}/k_BT}}{\int_0^W p_p(x)dx} \tag{3.49}$$

以 $p_p(x)\simeq N_B(x)$代入上面二式就得到小注入条件下基区电子分布 $n_p(x)$和基区电子电流 I_{En}的表示式。

若以 Q_G 表示单位面积中性基区的杂质总量(Q_G 称为 Gummel 数),即

$$Q_G=\int_0^W N_B(x)dx \tag{3.50}$$

则对于均匀掺杂(N_B 与 x 无关)的基区显然有 $Q_G=WN_B$。不难看出,只要在前面有关的公式中以 Q_G 代替 WN_B,就可以从均匀基区过渡到缓变基区的情形。

2. *基区扩展电阻和发射极电流集边效应*

图 3.9 是一个 n-p-n 晶体管的截面图,基极电流 I_B 从两个接触处流向发射区中央,在发射区的正下方与结面平行,与之相应的电阻称为基区扩展电阻,常以 $r_{bb'}$表示,其上的横向电压为

$$V_{bb'}=r_{bb'}I_B \tag{3.51}$$

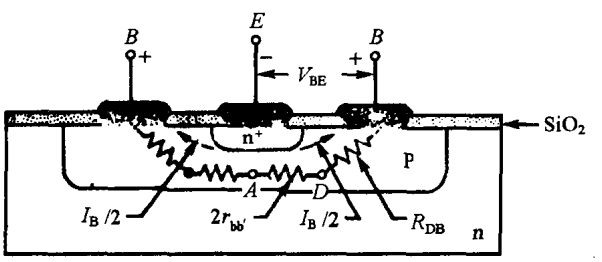


图 3.9 n-p-n 晶体管的基极电阻

$V_{bb'}$将明显改变作用在发射结势垒上的电压,使注入电流密度从边缘至中央指数下降。在 $V_{bb'}$远大于 k_BT/q 的情况下,发射极电流集中在发射结边缘附近,这种现象称为电流集边效应。

可以估算电流集边的有效宽度 ΔS_{eff},标准是其上的横向电压等于 k_BT/q,这时发射极电流密度已下降为边缘处的 1/e,对图 3.9 所示的结构,有

$$\Delta S_{eff}\simeq\frac{k_BT\mu_pN_BWl_E\beta_0}{I_C} \tag{3.52}$$

l_E 是发射极条的长度。上式虽然是粗略估算的结果，但从它可以看出，大电流时容易发生集边效应；同时，为了减弱集边效应，基区的掺杂浓度不能太低。

由于集边效应，发射极电流基本上同发射区的周长成正比，而不是同它的面积本身成正比，所以，降低发射极电流集边效应最有效的方法是采用周长/面积比很高的梳状结构。

3. 基区宽度调制

反偏 p-n 结的耗尽区宽度明显依赖于电压。当双极型晶体管工作在放大状态时，改变基极-集电极偏压将引起集电结耗尽区宽度变化，因此也引起中性基区宽度 W 的变化，这种现象称为基区宽度调制，也称 Early 效应。

基区宽度调制影响器件特性的表现之一是集电极电流 I_C 随偏压 V_{BC}变化，我们就由此出发来讨论这种效应。利用式(3.49)，集电极电流可表示为

$$I_C \simeq I_{En} = -\frac{qAD_n n_i^2 e^{qV_{BE}/k_BT}}{\int_0^W p_p(x)dx} \tag{3.53}$$

注意到 W 是 V_{BC}的函数，于是得到

$$\frac{\partial I_C}{\partial V_{BC}} = -I_C p_p(W)\left[\frac{1}{\int_0^W p_p(x)dx}\cdot\frac{\partial W}{\partial V_{BC}}\right] = -I_C/V_A \tag{3.54}$$

V_A 称为 Early 电压，定义为

$$V_A \equiv \frac{\int_0^W p_p(x)dx}{p_p(W)\dfrac{\partial W}{\partial V_{BC}}} \tag{3.55}$$

基区掺杂浓度越低，$\partial W/\partial V_{BC}$越大，$V_A$ 越小，Early 效应越显著。

实际上，V_A 与 V_{BC}近似无关，因此 V_A 常用 $V_{BC}=0$ 时的值来近似。根据这一条件以及式(3.54)，可以通过在 $V_{BC}=0$ 附近画出 $I_C \sim V_{CE}$曲线(共发射极输出特性曲线)的切线来确定 V_A，如图 3.10 所示。

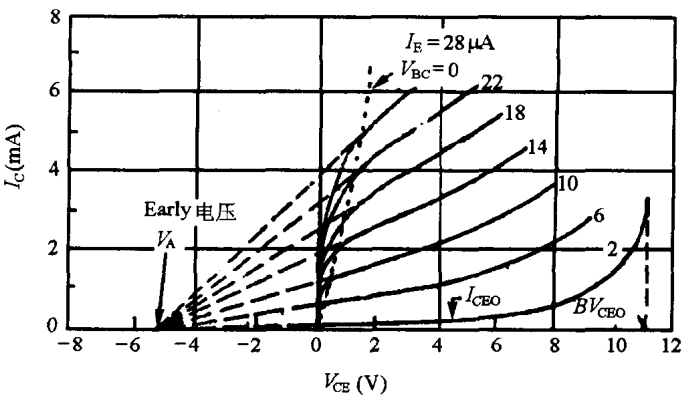


图 3.10　由共发射极输出特性曲线确定 V_A

以 Q_B' 表示基区中的多子电荷总量，则

$$Q_B' = qA\int_0^W p_p(x)\mathrm{d}x \tag{3.56}$$

$$\frac{\mathrm{d}Q_B'}{\mathrm{d}V_{BC}} = qAp_p(W)\frac{\mathrm{d}W}{\mathrm{d}V_{BC}} = C_{jc} \tag{3.57}$$

C_{jc}是集电结的小信号电容，将以上二式同式(3.55)比较，V_A 又可表示成

$$V_A = Q_B'/C_{jc} \tag{3.58}$$

大体说来，当 V_{BC}使基区宽度从 W_0 减小到 W_1 时，集电极电流 I_C 将按 W_0/W_1 的比例增大；同时，基区存储电荷量将减少，从而改变了晶体管的基极电流，电流增益 β_0 将随 W^{-2}变化。

4. 基区展宽效应

迄今一直假定，在放大状态下，$x = W$ 处的电子(少子)浓度为零，实际上只要有电子通过集电结，电子浓度至少等于 J_C/qv_s，J_C 是集电极电流密度，v_s 是电子饱和漂移速度。这时电子浓度在中性基区内被空穴中和，但在耗尽区内将改变正、负电荷层的浓度，负电荷层的浓度增加，正电荷层的浓度减小。若维持偏压 V_{BC}不变，则负电荷层必然变窄，正电荷层变宽，即中性基区倾向于加宽，整个耗尽区向 n^+衬底移动。在电子浓度 J_C/qv_s 超过集电区杂质浓度 N_C 的情形下，负电荷层将在集电区内形成，其电荷浓度为 $q(J_C/qv_s - N_C)$，而正电荷层在 n^+衬底内形成，也就是说中性基区宽度超过扩散时形成的原始基区宽度 W_B。这种现象称为基区展宽效应，或 Kirk 效应。

图 3.11 表示基区展宽时集电区的电场分布，这时最大场强从通常的基区-集电区界面($x=0$)移位到了集电区-衬底界面($x = W_C$)。J_{C0}表示基区展宽的临界电流密度，与之相应，泊松方程为

$$\frac{\mathrm{d}\mathscr{E}}{\mathrm{d}x} = -\frac{q(J_{C0}/qv_s - N_C)}{\varepsilon_s} \tag{3.59}$$

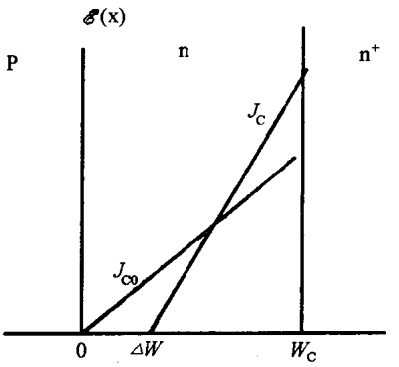


图 3.11　基区展宽时集电区内的电场分布

用单边突变近似并忽略内建电压，则有

$$J_{C0} = qv_s\left[\frac{2\varepsilon_s \mid V_{BC} \mid}{qW_C^2} - N_C\right] \tag{3.60}$$

$J_C > J_{C0}$时，基区向衬底方向扩展 ΔW，为

$$\Delta W = W_C\left[1 - \left(\frac{J_{C0} - qN_Cv_s}{J_C - qN_Cv_s}\right)^{\frac{1}{2}}\right] \tag{3.61}$$

基区展宽是集电极电流密度很大时出现的一种效应，集电区掺杂浓度较低时尤其容易发生。这种效应增加了基区的存储电荷，使电流增益 β_0 下降，并有损器件的频率响应。

5. 饱和电流和击穿电压

饱和电流 I_{CBO}和 I_{CEO}分别表示发射极开路(即 $I_E=0$)和基极开路(即 $I_B=0$)时的集电极电流，I_{CBO}比 I_{CEO}小得多，也比通常反向偏置 p-n 结(相当于 $V_{BE}=0$ 的情形)的饱和电流要小。这可以解释如下：集电极电流基本上由 $x=W$ 处的电子(少数载流子)的浓度梯度 dn_p/dx 决定，$I_E=0$ 时，$x=0$ 处的电子浓度梯度等于零，使 $x=W$ 处 dn_p/dx 很小，如图3.12所示；而 $I_B=0$ 时，在 V_{CE}一定的条件下，发射结受到轻微的正向偏置，从而使 $x=W$ 处 dn_p/dx 较大(见图3.12)。

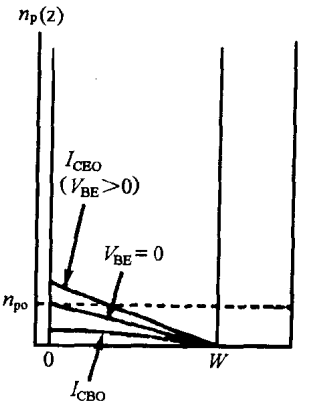


图 3.12 I_{CBO}和 I_{CEO}对应的基区电子(少子)分布

击穿电压 放大状态下，当 V_{CB}(共基极接法中的集电极-基极偏压)或 V_{CE}(共发射极接法中的集电极-发射极偏压)超过称之为击穿电压的临界值时，晶体管的集电极电流 I_C 急剧增加，这种现象称为雪崩击穿，其原因是集电结耗尽区内的电场太强而产生大量的电子和空穴(雪崩倍增)。

对于共基极接法，当集电结反向偏压增大到所谓击穿电压时，集电极电流迅速增大，其原因通常是雪崩击穿。以 BV_{CBO}表示发射极开路时的击穿电压，则该电压类似于 2.6 节讨论过的击穿电压，对于集电区掺杂远低于基区的情形有

$$BV_{CBO} \simeq \frac{\varepsilon_s \mathscr{E}_C^2}{2qN_c} \tag{3.62}$$

$\mathscr{E}_C$ 是临界击穿电场，N_C 是集电区掺杂浓度。

对于共发射极接法，基区开路($I_B=0$)时，$I_E=I_C=I$，I_{CBO}和 $\alpha_0 I_E$ 流过集电结时都增大 M 倍，即

$$M(I_{CBO} + \alpha_0 I) = I \tag{3.63}$$

或

$$I = \frac{MI_{CBO}}{1 - \alpha_0 M} \tag{3.64}$$

M 是集电结的倍增因子，可用经验公式表示为

$$M = \frac{1}{1 - (V/BV_{CBO})^n} \tag{3.65}$$

硅的 n 在 2 至 6 之间。应用击穿条件 $\alpha_0 M=1$ 和式(3.65)，得到共发射击穿电压为

$$BV_{CEO} = BV_{CBO}(1 - \alpha_0)^{1/n} \tag{3.66}$$

BV_{CEO}远低于 BV_{CBO}，是因为在共发射极的情形下发射结有注入作用。

击穿电压是对器件使用电压的限制。对于通过两次扩散(或离子注入)制作的双极晶体管，要提高击穿电压就必须降低集电区掺杂浓度，并保证集电区有足够的厚度。

基区穿通 在基区宽度很小或基区掺杂浓度很低的情形下，有可能在集电结发生雪崩击穿之前，集电结的耗尽区已经扩展到同发射结的耗尽区会合，即中性基区宽度 W 下降到零，这就是基区穿通，相应的集电结电压称为穿通电压，常以 V_{PT}表示。一旦基区穿通，发射区的电子直接受集电结的电场作用，相当于集电极与发射极短路，电流迅速增大。

设基区和集电压都是均匀掺杂，则根据第二章中的公式(2.20)，集电结耗尽区在基区内所占宽度为

$$d = \left[\frac{2\varepsilon_s}{q} \cdot \frac{N_C}{N_B(N_B + N_C)}(V_{CB} + V_{bi})\right]^{1/2} \tag{3.67}$$

$d = W_B$ 时基区穿通，相应的 V_{CB}等于 V_{pT}。忽略集电结内建电压 V_{bi}，有

$$V_{pT} \simeq \frac{q}{2\varepsilon_s}\frac{N_B}{N_C}(N_B + N_C)W_B^2 \tag{3.68}$$

$$\simeq \frac{q}{2\varepsilon_s}\frac{N_B^2}{N_C}W_B^2,\ N_C \ll N_B \tag{3.69}$$

由式(3.69)可见，对于一定的冶金基区宽度 W_B，只有基区掺杂浓度 N_B 较大时才能防止基区穿通，使器件的电压只受集电结耗尽区的雪崩倍增作用限制。

6. 高、低发射结偏压效应

发射结偏压 V_{BE}十分小时，其耗尽区的复合电流 I_{ER}在基极电流中占主导地位，它随 $\exp(qV_{BE}/mk_BT)$变化($m \simeq 2$)；而集电极电流主要来自注入基区的电子电流，大致随 $\exp(qV_{BE}/k_BT)$变化。因此 β_0 大致随 $\exp\left[\left(1-\frac{1}{m}\right)qV_{BE}/k_BT\right]$变化。这种效应称为 Sah 效应。当 V_{BE} 增大到发生大注入现象时，相当于增加了基区的掺杂浓度(基区电导调制效应)，使发射效率 γ 降低，所以 β_0 下降，这种效应称为 Webster 效应。

如果将 $\beta_0 = I_{Cn}/(I_{EP} + I_{ER} + I_{BR})$改写如下：

$$\beta_0^{-1} = \frac{I_{EP} + I_{ER} + I_{BR}}{I_{Cn}} \simeq \frac{I_{EP}}{I_{En}} + \frac{I_{ER}}{I_{En}} + \frac{I_{BR}}{I_{En}} \tag{3.70}$$

则在对它进行逐项分析以后将得到

$$\beta_0^{-1} = \alpha_1 I_C^{\left(\frac{1}{m}-1\right)} + \alpha_2 + \alpha_3 I_C \tag{3.71}$$

式中 α_1，α_2 和 α_3 与 I_C 无关，只决定于器件的工艺参数，在小电流下，$\beta_0 \propto I_C^{(1-1/m)}$，随 I_C 的增大而上升；在中等电流下，β_0 与 I_C 无关；在大电流下，$\beta_0 \propto 1/I_C$，随 I_C 的增大而下降。图 3.13(a)画出了 β_0 随 I_C 变化的典型曲线，图 3.13(b)则画出了集电极电流 I_C、基极电流 I_B 和发射结偏压 V_{BE}的关系。图中电流用对数坐标，所以同一 V_{BE}下 I_C 和 I_B 差别实际上说明 β_0 的大小，而差别大小随 V_{BE}的变化反映了 β_0 随 V_{BE}的变化。

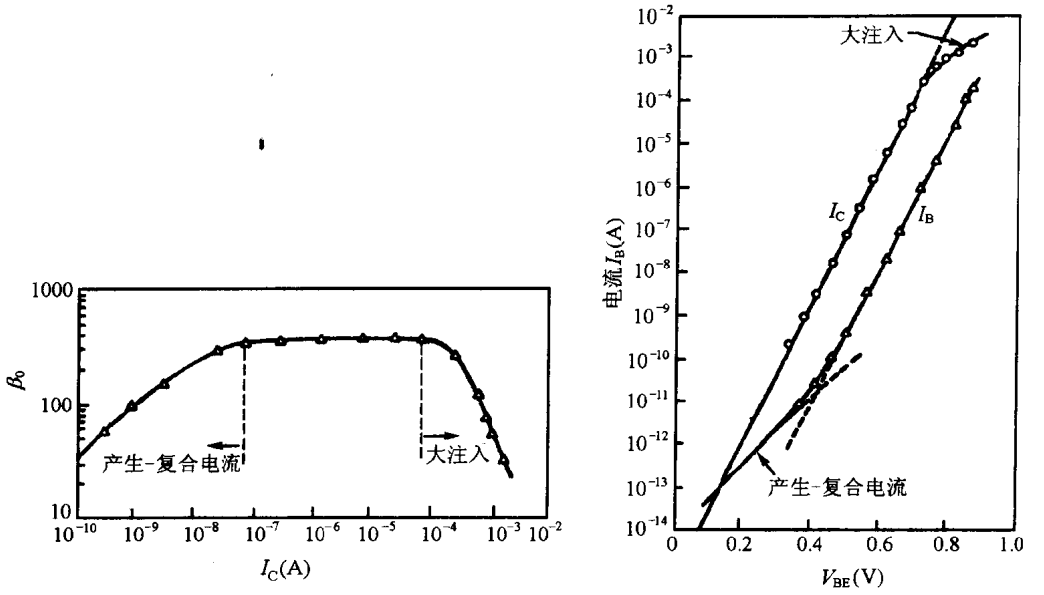
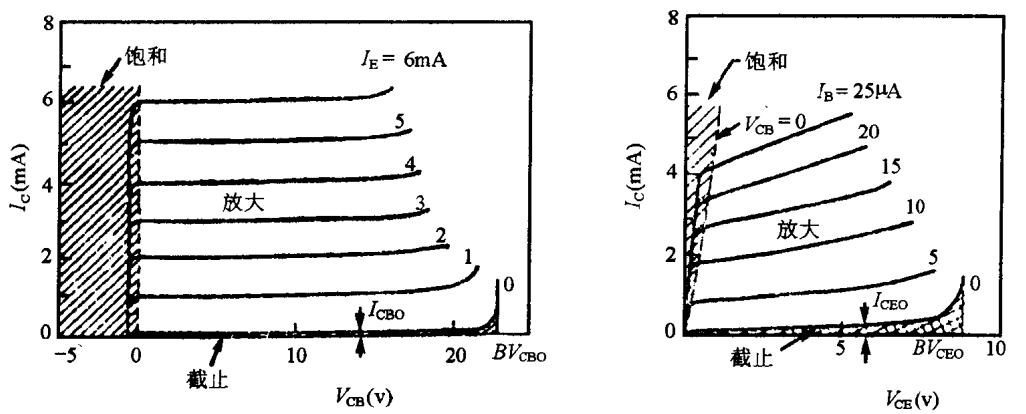


图 3.13 (a) 晶体管共发射极电流增益 β_0 与集电极电流 I_c 的关系；(b) 集电极电流 I_c、基极电流 I_B 与发射结电压 V_{BE}的关系

3.2.5 输出特性

图 3.14 画出了 n-p-n 晶体管的一组共基极接法和共发射极接法的典型的输出特性曲线。



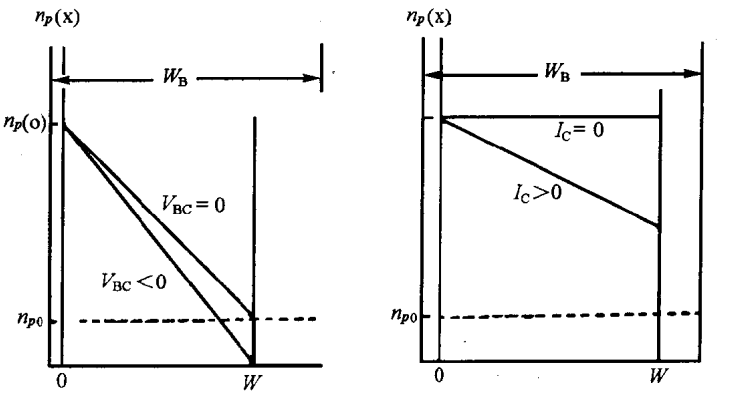
(a) 共基极接法　　(b) 共发射极接法

图 3.14 n-p-n 晶体管的输出特性

对于共基极接法，由图 3.14(a)看到，集电极电流基本上等于发射极电流(即 $\alpha_0 \approx 1$)，实际上不依赖于 V_{CB}。这和式(3.9)给出的理想晶体管的性能非常一致。集电极电流实际上保持

为常量，甚至在 V_{CB}降到零时，集电极仍然抽取电子。图 3.15(a)所示的电子分布表明了这一点。因为从 $V_{CB}>0$ 变到 $V_{CB}=0$ 时 $x=W$ 处的电子浓度只有微小的变化，所以在整个放大工作状态范围内，集电极电流基本保持不变。为了使集电极电流减小到零，我们必须在基区-集电区结上加一小的正向偏压(即 $V_{CB}<0$)使其处于饱和状态，如图 3.15(b)所示。这个正向偏压(对硅，$V_{CB}\sim -1V$)将足以增加 $x=W$ 处的电子浓度，使之等于发射区在 $x=0$ 处的电子浓度，这样，使 $x=W$ 处的电子浓度梯度减小，从而使集电极电流下降到零。当 V_{CB}增大到某一数值(发射极开路时为 BV_{CBO})时，集电极电流迅速增加，通常这是集电结的雪崩击穿所致。当基区宽度很小，或基区掺杂浓度很低，击穿也可能是基区穿通所致。I_{CBO}是发射极开路($I_E=0$)时的集电极电流，如以前所论述的，$x=0$ 处的电子浓度梯度是零，降低了 $x=W$ 处的电子浓度梯度，所以 I_{CBO}应小于发射结短路($V_{BE}=0$)时的电流。

对于共发射极接法，I_B 一定(即 V_{BE}一定)时，随 V_{CE}增大(即 V_{CB}增大，亦即中性基区宽度 W 减小)，由于 Early 效应，β_0(随 W^{-2}变化)增大，所以集电极电流随 V_{CE}的增加而增加，曲线明显倾斜。由 $V_{CE}=V_{CB}+V_{BE}$可知，V_{CE}在两个结之间分压，为了维持 I_B 一定，V_{BE}实际保持恒定，所以当 V_{CE}减小到某一值时将使 $V_{CB}=0$，晶体管进入饱和状态，I_C 开始下降，在 $V_{CE}=0$ 时下降到零。如前所述，由于发射结的注入作用，击穿电压 BV_{CEO}比 BV_{CBO}低得多，饱和电流 I_{CEO}比 I_{CBO}大得多。



(a) $V_{CB}=0$ 或 $V_{CB}>0$ 的放大状态　　(b) 两个结都处于正向偏置的饱和状态

图 3.15　n-p-n 晶体管基区内电子(少数载流子)浓度分布

3.3　双极型晶体管模型

提出过多种模型来概括双极型晶体管的电学特性，其中得到广泛使用的是 Ebers-Moll(E-M)模型和 Gummel-Poon(G-P)模型。E-M 模型使器件的电学特性和器件的工艺参数相联系，而 G-P 模型则建立在器件电学特性和基区多子电荷相联系的基础之上。

3.3.1　E-M 模型

这一模型包括两个二极管和两个电流源。对于一个 n-p-n 晶体管，基区-发射区结和基区-集电区结用具有共同 p 区的两个背靠背的二极管描述，二极管电流分别为

$$I_F = I_{F0}(e^{qV_{BE}/k_BT} - 1) \tag{3.72}$$

$$I_R = I_{R0}(e^{qV_{BC}/k_BT} - 1) \tag{3.73}$$

I_{F0}和 I_{R0}是通常二极管的饱和电流。电流源则代表一个结的注入而通过另一个结的电流，分别以 $\alpha_F I_F$ 和 $\alpha_R I_R$ 表示，前者与基区-集电区结二极管并联。后者与基区-发射区结二极管并联，α_F 和 α_R 分别称为正向和反向共基极电流增益。于是，一个理想的 n-p-n 晶体管可用图3.16所示的等效电路表示。

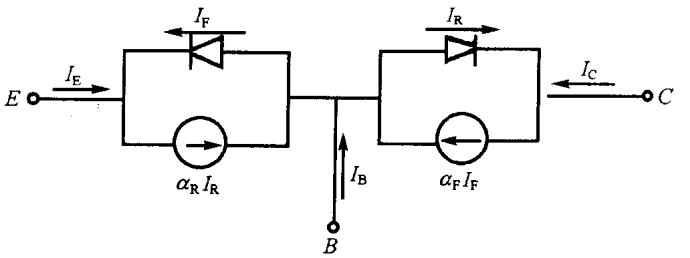


图 3.16　n-p-n 晶体管基本 E-M 模型的等效电路

由图 3.16，端电流可表示为

$$I_E = -I_F + \alpha_R I_R \tag{3.74}$$

$$I_C = \alpha_F I_F - I_R \tag{3.75}$$

$$I_B = (1 - \alpha_F) I_F + (1 - \alpha_R) I_R \tag{3.76}$$

将式(3.72)—(3.75)联立，可得出基本的 E-M 模型方程：

$$I_E = -I_{F0}(e^{qV_{BE}/k_BT} - 1) + \alpha_R I_{RO}(e^{qV_{BC}/k_BT} - 1) \tag{3.77}$$

$$I_C = \alpha_F I_{F0}(e^{qV_{BE}/k_BT} - 1) - I_{RO}(e^{qV_{BC}/k_BT} - 1) \tag{3.78}$$

利用以上二式，不难证明下述关系：

$$I_{F0} = \frac{I_{EBO}}{1 - \alpha_F \alpha_R}, \quad I_{R0} = \frac{I_{CBO}}{1 - \alpha_F \alpha_R} \tag{3.79}$$

式中，I_{EBO}和 I_{CBO}分别为集电极开路($I_C = 0$)时的发射极饱和电流和发射极开路($I_E = 0$)时的集电极饱和电流。

将式(3.77)，(3.78)分别和式(3.27)，(3.28)比较，有

$$\begin{aligned} I_{FO} &= -a_{11} \\ \alpha_R I_{RO} &= a_{12} \\ \alpha_F I_{FO} &= a_{21} \\ I_{RO} &= -a_{22} \end{aligned} \tag{3.80}$$

式(3.80)使基本的 E-M 模型的参数 I_{FO}，I_{RO}，α_F 和 α_R 通过 a_{11}，a_{12}，a_{21}和 a_{22}同器件的工艺参数相联系。由于二端口网络器件的互易性质，有 $a_{12} = a_{21}$，即 $\alpha_R I_{RO} = \alpha_F I_{FO}$，四个模型参数中只有三个是独立的。

以上讨论的是基本的 E-M 模型，它只是一种非线性直流模型，为了和后来改进的各种模型相区别，通常将它记为 EM_1 模型。在 EM_1 的基础上计及非线性电荷存储效应和欧姆电阻，就构成第二级复杂程度的 EM_2 模型。第三级复杂程度的 EM_3 模型则还包括各种二级效应，

如基区宽度调制，基区展宽效应以及器件参数随温度的变化等。EM_3 模型和下面将要讨论的 G-P 模型是等价的。

3.3.2 G-P 模型

模型的数学表示 半导体中的电流密度可表示为

$$J_n = \mu_n n \frac{dE_{Fn}}{dx} \tag{3.81}$$

$$J_p = \mu_p p \frac{dE_{Fp}}{dx} \tag{3.82}$$

E_{Fn}, E_{Fp}分别为电子、空穴的准费米能级。对于 n-p-n 晶体管，由于基区中不存在明显的空穴电流，可假定 $dE_{Fp}/dx = 0$。由此可以得到

$$J_n \approx \mu_n n \frac{d}{dx}(E_{Fn} - E_{Fp}) \tag{3.83}$$

应用关系式 $np = n_i^2 \exp[(E_{Fn} - E_{Fp})/k_B T]$及 $D_n = \frac{k_B T}{q}\mu_n$，上式可改写为

$$J_n = qD_n \frac{1}{p}\frac{d}{dx}(np) \tag{3.84}$$

为了下面的讨论方便，现在转用 x_E 和 x_C 来表示基区边界的位置。在忽略基区复合和假定 D_n 为常量(否则以 $\overline{D}_n$ 代替)的条件下，将上式从 x_E 至 x_C 积分，可得

$$J_n = \frac{-qD_n n_i^2[e^{qV_{BE}/k_B T} - e^{qV_{BC}/k_B T}]}{\int_{X_E}^{X_C} p(x)dx} \tag{3.85}$$

我们采用从发射极向集电极的传输电流 I_{CC}，将上式改写为

$$\begin{aligned} I_{CC} &= \frac{A^2 q^2 D_n n_i^2}{qA\int_{X_E}^{X_C} p(x)dx}[(e^{qV_{BE}/k_B T} - 1) - (e^{qV_{BC}/k_B T} - 1)] \\ &= \frac{I_S Q_{B0}}{Q_B}[(e^{qV_{BE}/k_B T} - 1) - (e^{qV_{BC}/k_B T} - 1)] \\ &= I_{CE} - I_{EC} \end{aligned} \tag{3.86}$$

式中 I_{CE}称为正向传输电流，I_{EC}称为反向传输电流，$I_S = A^2 q^2 D_n n_i^2 / Q_{B0}$为晶体管的饱和电流。

基区多子电荷 G-P 模型的主要特点是把双极型晶体管的电学性质和基区多子电荷的积分形式联系在一起，所以问题转化为基区的电荷模型，表示成

$$Q_B = qA\int_{X_E}^{X_C} p(x)dx \tag{3.87}$$

它由 5 个分量组成：

$$Q_B = Q_{B0} + Q_{jE} + Q_{jC} + Q_{dE} + Q_{dC} \tag{3.88}$$

(1) $Q_{B0} = qA\int_{X_{E0}}^{X_{C0}} N_B(x)dx$

Q_{B0}代表热平衡($V_{BE} = V_{BC} = 0$)时基区的多子电荷总量。有外加电压时，基区边界将发生移动，造成基区内的电荷存储，这些附加电荷从物理上可区分为其后所述的 4 个分量。

(2) $Q_{jE} = -qA\int_{X_{E0}}^{X_E} N_B(x)\mathrm{d}x$

Q_{jE}代表发射结正偏时其耗尽区宽度变化而使基区多子电荷增加的数量，故它又可通过发射结电容 C_{jE}表示为

$$Q_{jE} = qA\int_0^{V_{BE}} C_{jE}\mathrm{d}V_{BE} \tag{3.89}$$

引入

$$\overline{C}_{jE} = \frac{1}{V_{BE}}\int_0^{V_{BE}} C_{jE}\mathrm{d}V_{BE} \tag{3.90}$$

则

$$Q_{jE} = \overline{C}_{jE}V_{BE} = \frac{Q_{B0}V_{BE}}{V_A'} \tag{3.91}$$

$$V_A' = \frac{Q_{B0}}{\frac{1}{V_{BE}}\int_0^{V_{BE}} C_{jE}\mathrm{d}V_{BE}} = \frac{Q_{B0}}{\overline{C}_{jE}} \tag{3.92}$$

V_A'称为逆向 Early 电压。习惯上把集电结电压引起的基区宽度变化称为 Early 效应。

(3) $Q_{jC} = -qA\int_{X_C}^{X_{C0}} N_B(x)\mathrm{d}x$

Q_{jC}代表集电结正偏时其耗尽区宽度变化而使基区多子电荷增加的数量。它又可通过集电结容 C_{jC}表示成

$$Q_{jC} = \int_0^{X_{BC}} C_{jC}\mathrm{d}V_{BC} \tag{3.93}$$

或

$$Q_{jC} = \frac{Q_{B0}V_{BC}}{V_A} \tag{3.94}$$

V_A称为 Early 电压：

$$V_A = \frac{Q_{B0}}{\frac{1}{V_{BC}}\int_0^{V_{BC}} C_{jC}\mathrm{d}V_{BC}} = \frac{Q_{BC}}{\overline{C}_{jC}} \tag{3.95}$$

(4) $Q_{dE} + Q_{dC} = qA\int_{X_E}^{X_C}(p - N_B)\mathrm{d}x$

代表基区中存储电荷的数量。由于电子和空穴成对存储，Q_{dE}和 Q_{dC}为注入存储使基区多子电荷增加的数量。其中，Q_{dE}代表发射结正偏、集电结零偏时基区中多子电荷的增加量：

$$Q_{dE} = qA\int_{X_E}^{X_{C0}}(p - N_B)\mathrm{d}x \tag{3.96}$$

可以把这部分电荷同正向传输电流 I_{CE}联系起来：

$$Q_{dE} = B\tau_B I_{CE} \tag{3.97}$$

τ_B为基区渡越时间；B为基区展宽系数，小注入时 $B=1$，大注入时 $B>1$，Q_{dC}代表集电结正偏，反射结零偏时基区中多子电荷的增加量，它同反向传输电流 I_{EC}的关系为

$$Q_{dC} = \tau_{BR}I_{EC} \tag{3.98}$$

τ_{BR}称为反向基区渡越时间。

在 Q_B 的五个量中，Q_{B0}，Q_{jE}，Q_{jC}3 个分量是和基区杂质分布有关的多子电荷量，Q_{dE}，Q_{dC}两个分量是由于注入存储在基区中的多子电荷量。小注入时，注入的电荷和掺杂相比是可以忽略的，所以前 3 个分量是 Q_B 的主要分量，只有大注入时后 2 个分量才变得显著。

总结上述分析，Q_B 可以表示成

$$Q_B = Q_{B0} + Q_{jE} + Q_{jC} + B\tau_B I_{CE} + \tau_{BR} I_{EC}$$

$$= Q_{B0} + Q_{jE} + Q_{jC} + \frac{Q_{B0} I_S}{Q_B}[B\tau_B(e^{qV_{BE}/k_BT} - 1) + \tau_{BR}(e^{qV_{BC}/k_BT} - 1)]$$

令

$$Q_{BO} + Q_{jE} + Q_{jC} = Q_1$$

$$Q_{B0} I_S[B\tau_B(e^{qV_{BE}/k_BT} - 1) + \tau_{BR}(e^{qV_{BC}/k_BT} - 1)] = Q_2$$

则有

$$Q_B = Q_1 + Q_2/Q_B \tag{3.99}$$

解之得

$$Q_B = \frac{Q_1}{2} + \sqrt{\frac{Q_1^2}{4} + Q_2} \tag{3.100}$$

作为例子，应用式(3.100)来分析双极型晶体管的放大状态。很容易得到，小注入时有

$$Q_B \approx Q_1$$

$$I_C \approx I_{CE} = \frac{Q_{B0} I_s}{Q_1}(e^{qV_{BE}/k_BT} - 1)$$

大注入时，有

$$Q_B \approx \sqrt{Q_2} \approx \sqrt{Q_{B0} I_s B\tau_B} e^{qV_{BE}/2k_BT}$$

$$I_C \approx \sqrt{\frac{Q_{B0} I_s}{B\tau_B}} e^{qV_{BE}/2k_BT}$$

等效电路和端电流 计入欧姆电阻 r_E，r_B 和 r_C，以及产生-复合电流 I_{ER}和 I_{CR}后，n-p-n 晶体管的等效电路如图 3.17 所示，图中 I_{ER}和 I_{CR}可分别表示为

$$I_{ER} = I_1(e^{qV_{B'E'}/m_E k_B T} - 1) \tag{3.101}$$

$$I_{CR} = I_2(e^{qV_{B'C'}/m_C k_B T} - 1) \tag{3.102}$$

I_1，I_2，m_E 和 m_C 是通过实验来确定的参量，不受外加电压直接控制。于是，基极电流为

$$I_B = \frac{dQ_B}{dt} + \frac{I_{CE}}{\beta_F} + \frac{I_{EC}}{\beta_R} + I_{ER} + I_{CR} \tag{3.103}$$

式中，$\frac{dQ_B}{dt}$是提供基区多子增量的电流，包括对结电容 C_{jE}，C_{jC}和扩散电容 C_{dE}，C_{dC}的充电；集电极电流为

$$I_C = I_{CC} - I_{CR} - \frac{I_{EC}}{\beta_R} - C_{jC}\frac{dV_{B'C'}}{dt} - \tau_{BR}\frac{dI_{EC}}{dt} \tag{3.104}$$

发射极电流为

$$I_E = -I_{CC} - I_{ER} - \frac{I_{CE}}{\beta_F} - C_{jE}\frac{dV_{B'E'}}{dt} - B\tau_B\frac{dI_{CE}}{dt} \tag{3.105}$$

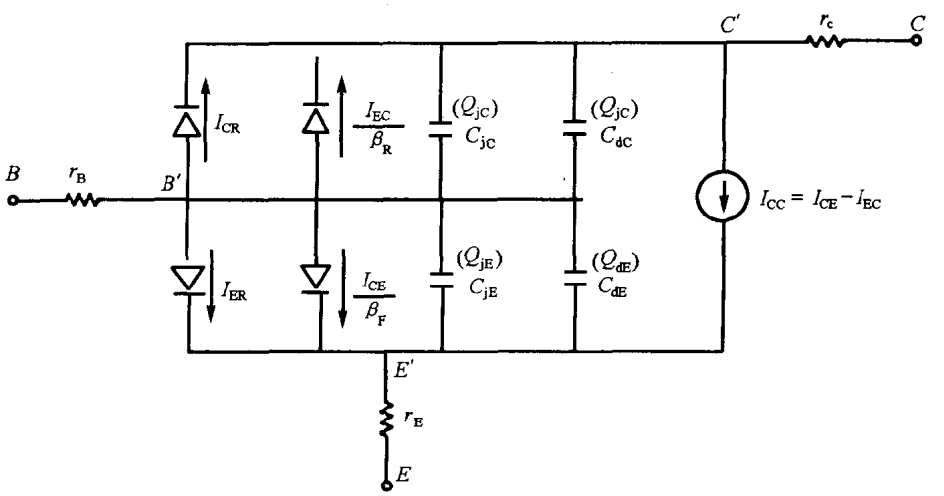


图 3.17　n-p-n 晶体管 G-P 模型的电路图

3.4　双极型晶体管的频率特性

3.4.1　小信号交流等效电路

在小信号(信号电压小于 k_BT/q)条件下,可以用线性等效电路来表示晶体管,图 3.18 表示共发射极接法的小信号等效电路。图中,$C_E = C_{jE} + C_{dE}$是发射结电容,包括耗尽层电容 C_{jE} 和扩散电容 C_{dE};C_{jC}是集电结耗尽层电容;r_B 和 r_C 分别为基极和集电极的串联电阻(发射极串联电阻 r_E 很小而被忽略);g_m 称为跨导,利用放大状态下 $I_C \approx -\alpha_F I_{FO} e^{qV_{BE}/k_BT}$,可得

$$g_m \equiv \frac{\partial I_C}{\partial V_{BE}} = \frac{qI_C}{k_BT} \tag{3.106}$$

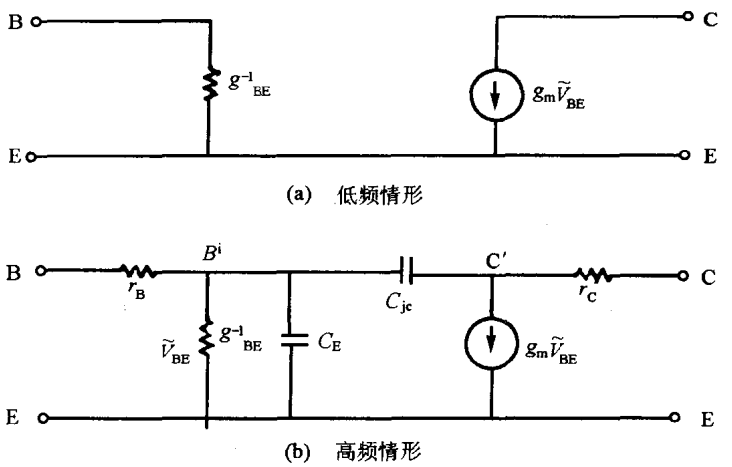


图 3.18　双极型晶体管小信号等效电路

g_{BE}称为输入电导，定义为

$$g_{BE} \equiv \frac{\partial I_B}{\partial V_{BE}} = \frac{\partial (I_C/\beta_0)}{\partial V_{BE}} = g_m/\beta_0 \tag{3.107}$$

显然，如果把上面二式中的电流和电压的增量看作交流量，则由它们就得到低频情形下的等效电路。

3.4.2 电流增益随频率的变化

1. β 随工作频率的变化

β 表示共发射极交流短路电流增益。在输出端交流短路时，图 3.18(b)可改画为图 3.19，图中 $C_\mu = (1 + r_C g_m) C_{jC}$是包括 C_{jC}和 r_C 作用的等效电容。

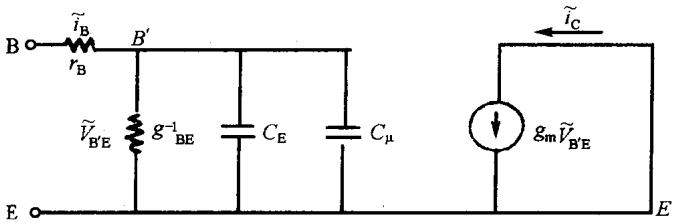


图 3.19 输出端交流短路时双极型晶体管的高频小信号电路

由图 3.19 可得

$$\beta \equiv \frac{\tilde{i}_C}{\tilde{i}_B} = \frac{g_m}{g_{B'E} + j\omega(C_E + C_\mu)} = \frac{\beta_0}{1 + j\omega/\omega_\beta} \tag{3.108}$$

其中

$$\omega_\beta = \frac{g_{B'E}}{C_E + C_\mu} \tag{3.109}$$

式(3.108)表示了 β 随频率的变化，通常写成

$$\beta = \frac{\beta_0}{1 + jf/f_\beta} \tag{3.110}$$

$$|\beta| = \beta_0 / \sqrt{1 + (f/f_\beta)^2} \tag{3.111}$$

2. α 随工作频率的变化

α 表示共基极交流短路电流增益，即输出端交流短路时交流电流$\tilde{i}_C$与$\tilde{i}_E$之比。在交流情形下，仍有

$$\beta = \frac{\alpha}{1 - \alpha} \tag{3.112}$$

由式(3.112)及(3.110)可得

$$\alpha = \frac{\alpha_0}{1 + jf/f_\alpha} \tag{3.113}$$

$$f_\alpha = (\beta_0 + 1) f_\beta \tag{3.114}$$

电流增益随工作频率的变化关系表示在图 3.20 中。电流增益的数值常用分贝(dB)表示,分贝的数值是把 α 或 β 的数值取以 10 为底的对数再乘以 20,即 $20\lg|\alpha|$,$20\lg|\beta|$。

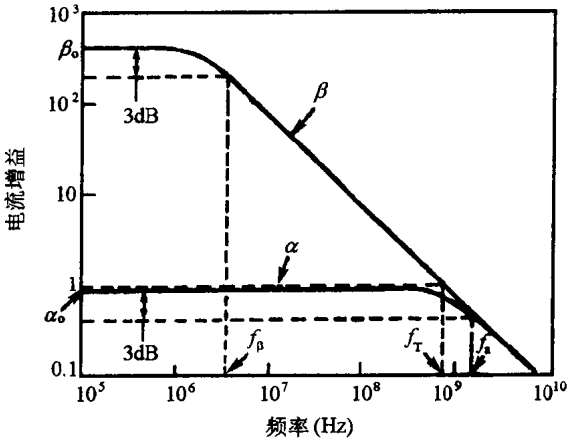


图 3.20 电流增益和工作频率的关系

3.4.3 频率参数

图 3.20 引入了 3 个频率参数,它们表征双极型晶体管的电流放大能力。f_α 称为共基极截止频率,定义为 $|\alpha|$ 下降到 $\alpha_0/\sqrt{2}$(下降 3dB)时的频率,f_β 称为共发射极截止频率,定义为 $|\beta|$ 下降到 $\beta_0/\sqrt{2}$ 时的频率,f_T 称为特征频率,定义为 $|\beta|$ 等于 1 时的频率。

由(3.109)式,得

$$f_\beta = \frac{1}{2\pi}\omega_\beta = \frac{1}{2\pi}\frac{g_{B'E}}{C_E + C_\mu} \tag{3.115}$$

将上式右方各量的表达式代入,即得

$$f_\beta = \frac{1}{2\pi}\cdot\frac{1}{\tau_F + \dfrac{k_B T}{qI_C}(C_{jE} + C_{jC}) + \gamma_C C_{jC}} \tag{3.116}$$

其中 τ_F 称为正向渡越时间,略大于基区渡越时间 τ_B,可理解为移走器件内(主要是基区和发射区)存储电荷所需时间。

在式(3.111)中令 $|\beta| = 1$ 和 $f = f_T$,得到

$$\begin{aligned} f_T &\simeq \beta_0 f_\beta \\ &= \frac{1}{2\pi}\left[\tau_F + \frac{k_B T}{qI_C}(C_{jE} + C_{jC}) + \gamma_C C_{jC}\right]^{-1} \end{aligned} \tag{3.117}$$

此外,从式(3.110)不难看出,$f \gg f_\beta$ 时有 $|\beta| f = \beta_0 f_\beta = f_T$,即频率和增益的乘积等于一个常数,这个常数就是器件的特征频率 f_T。为此,f_T 又称增益-带宽乘积,并说明可以在比 f_T 低得多的频率下测量 β 来得到 f_T。

在式(3.117)中加入电子渡越集电结耗尽区的渡越时间 $\tau_d = X_{dC}/v_s$(X_{dC}表示集电结耗尽

区宽度，v_s是电子的饱和漂移速度)，则 f_T 可重新表示为

$$f_T = (2\pi\tau_{EC})^{-1} \tag{3.118}$$

τ_{EC}是电子从发射极到集电极的有效渡越时间，可表示为

$$\tau_{EC} = \tau_E + \tau_B + \tau_D + \tau_C \tag{3.119}$$

其中 $\tau_E = (k_B T/qI_C)C_{jE}$是发射结充电时间，$\tau_B \simeq \tau_F$ 是电子渡越基区的时间，τ_D 是电子漂移渡越集电结耗尽区的时间，$\tau_C = (k_B T/qI_C + \gamma_C)C_{jC}$是集电结充电时间。要提高双极晶体管的最高工作频率，就要降低这些时间常数，主要是减小结电容和基区厚度。在早期的双极晶体管中，基区渡越时间 τ_B 是决定特征频率的主要参数。由于半导体微细加工技术的发展，现代双极晶体管的基区厚度可以做到很小，τ_B 的作用显著减弱，所以必须考虑结电容(特别是 C_{jC})的影响。如前所述，双极晶体管的基极电流横向流动，所以当器件的横向尺寸缩小时，基区电阻 r_B 因为电流流经的路程缩短而减小，结电容(C_{jE}和 C_{jC})因为结面积变小也将减小。如果晶体管的纵向尺寸和横向尺寸都按同一比例因子 α 缩小，则 r_B 不变，而结电容将减小到原来的 $1/\alpha$ 倍。集电结的面积通常比发射结的面积大得多，所以减小集电结面积是十分重要的，目前先进的双极晶体管(例如多晶硅发射极晶体管)技术常用深槽隔离方法来限制集电结面积。

共基极截止频率 f_α 可以通过式(3.114)由 f_β 给出。另一个频率参数也经常使用，这就是最高振荡频率 f_M，它被定义为双极晶体管的功率增益等于 1 时的频率，可表示为

$$f_M = \sqrt{\frac{f_T}{8\pi(r_B + \pi f_T L)C_{jC}}} \simeq f_T(8\pi r_B f_T C_{jC})^{-1/2} \tag{3.120}$$

式中 L 为发射极引线电感，r_B 为基极电阻。f_M 是双极晶体管放大能力的极限频率。

3.4.4 基区渡越时间

少数载流子渡越中性基区时间 τ_B 的延迟作用对于双极晶体管频率响应能力是一个重要限制，特别是基区宽度比较大时，τ_B 实际上决定着晶体管的特征频率。τ_B 可表示为

$$\tau_B = \int_0^W \frac{dx}{v(x)} \tag{3.121}$$

$v(x)$是基区内少数载流子的有效速度，它与电流的关系为

$$I_n = qn_p(x)v(x)A \tag{3.122}$$

故有

$$\tau_B = \int_0^W \frac{qAn_p(x)}{I_n(x)}dx \tag{3.123}$$

A 是晶体管横截面，$n_p(x)$是基区内电子(少子)在位置 x 上的浓度。忽略基区复合，则 $I_n = I_{En} \simeq I_C$，与 x 无关，这时有

$$\tau_B = \frac{qA\int_0^W n_p(x)dx}{I_C} \simeq \frac{Q_B}{I_C} \tag{3.124}$$

$Q_B \simeq qA\int_0^W n_p(x)dx$ 是基区内存储电荷总量。上式说明 τ_B 可理解为以电流 I_C 移走基区存储电荷所需要的时间。

对于均匀基区和小注入情形，由表示 Q_B 的式(3.37)和表示 I_C 的式(3.42)，有

$$\tau_B = \frac{W^2}{2D_n} \tag{3.125}$$

D_n 为电子的扩散系数。由于大多数半导体中电子的扩散系数比空穴的大，高频晶体管大都采用 n-p-n 结构。为了改进频率响应，必须缩短 τ_B，所以高频晶体管的基区宽度必须很小。在现代先进的半导体工艺水平下，基区宽度能够严格控制，降低结电容、引线电感和基极电阻在设计高频晶体管中将是十分重要的。

对于不限于均匀掺杂和小注入的一般情形，可用下式

$$I_C \simeq I_{En} = \frac{qAD_n n_i^2 e^{qV_{BE}/k_B T}}{\int_0^W p_p(x)dx} \tag{3.126}$$

有

$$\tau_B = \frac{\int_0^W p_p(x)dx \int_0^W n_p(x)dx}{D_n n_i^2 e^{qV_{BE}/k_B T}} \tag{3.127}$$

或者，将表示 $n_p(x)$ 的公式(3.48)代入式(3.123)，有

$$\tau_B = \frac{1}{D_n}\int_0^W \frac{1}{p_p(x)}\left[\int_x^W p_p(x')dx'\right]dx \tag{3.128}$$

引入 $y = x/W$，则式(3.128)变为

$$\begin{aligned}\tau_B &= \frac{W^2}{D_n}\int_0^1 \frac{1}{p_p(y)}\left[\int_y^1 p_p(x')dx'\right]dy \\ &= \frac{W^2}{\nu D_n}\end{aligned} \tag{3.129}$$

其中 ν 是上式中复式积分的倒数，即

$$\nu = \left\{\int_0^1 \frac{dy}{p_p(y)}\left[\int_y^1 p_p(x')dx'\right]\right\}^{-1} \tag{3.130}$$

对于扩散形成的非均匀基区，$\nu > 2$，基区内建电场的作用是使 τ_B 缩短。

3.4.5 功率-频率限制

双极晶体管的最大输出功率 P_m 受雪崩击穿电压 V_m(或临界电场 $\mathscr{E}_c$)和饱和漂移速度 v_s 的限制。双极晶体管的最大允许电压为

$$V_m = \mathscr{E}_c L_{EC} \tag{3.131}$$

L_{EC}是发射极和集电极之间的距离。假设载流子以饱和漂移速度运动，则它从发射极到集电极的渡越时间为 $\tau_{EC} = L_{EC}/v_s$，故晶体管的最高截止频率为

$$f_T = \frac{1}{2\pi\tau_{EC}} = \frac{v_s}{2\pi L_{EC}} \tag{3.132}$$

所以有

$$V_m f_T = \frac{\mathscr{E}_c v_s}{2\pi} \tag{3.133}$$

流过双极晶体管的最大允许电流可表示为

$$I_m = \frac{V_m}{\chi_C} \tag{3.134}$$

其中 χ_C 为电抗，可表示为

$$\chi_C = \frac{1}{\omega_T C_{jC}} = \frac{1}{2\pi f_T C_{jC}} \tag{3.135}$$

于是，

$$P_m = V_m^2/\chi_C \tag{3.136}$$

最终可得功率-频率极限满足下式

$$(P_m \chi_C)^{1/2} f_T = \frac{\mathscr{E}_c v_s}{2\pi} \tag{3.137}$$

对于硅，$\mathscr{E}_c \simeq 3\times10^5$v/cm，$v_s \simeq 6\times10^6$cm/s。上式说明，对于一定的器件阻抗，器件的功率输出能力随截止频率升高而下降。

3.5 双极型晶体管的开关特性

3.5.1 晶体管的开关作用

对于数字应用，双极型晶体管设计成一个开关。在这些应用中，用一个小的基极电流去改变集电极电流，使得在很短的时间内从高电压、低电流的关断状态，变化到低电压、大电流的导通状态；或者相反。基本开关电路如图 3.21(a)所示，其中发射结电压 V_{BE} 突然从负值变化到正值。晶体管的输出电流如图 3.21(b)所示。因为发射结和集电结都处于反向偏置，开始时集电极电流很小。该电流将沿负载线通过放大区，最后达到大电流水平，这时两个结都变成正向偏置。这样，在关断状态(相应于截止状态)下，晶体管发射极端和集电极端之间实际上是开路的；而在导通状态(相应于饱和状态)下，晶体管发射极和集电极之间基本上相当于短路。因此，以这种方式工作的晶体管几乎与一个理想开关没有差别。

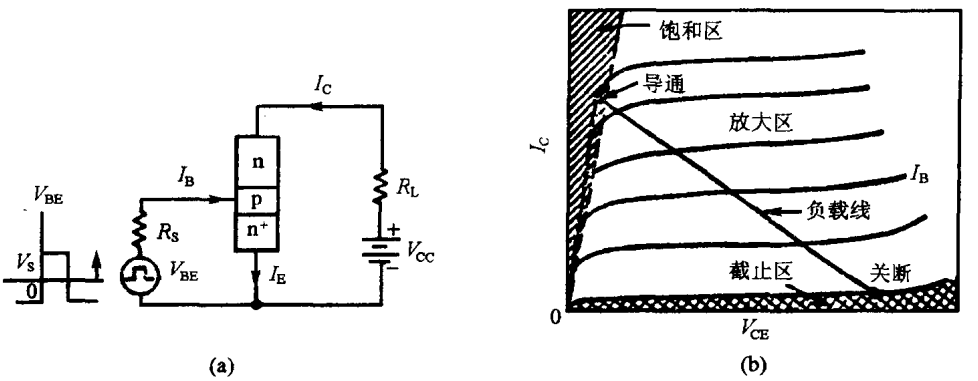


图 3.21 (a) 基本开关电路；(b) 晶体管的开关作用

3.5.2 关断和导通阻抗

我们用 E-M 模型来讨论双极型晶体管的开关阻抗。为了方便，将式(3.77)和(3.78)重写如下：

$$I_E = -I_{F0}(e^{qV_{BE}/k_BT} - 1) + \alpha_R I_{R0}(e^{qV_{BC}/k_BT} - 1) \tag{3.138a}$$

$$I_C = \alpha_F I_{F0}(e^{qV_{BE}/k_BT} - 1) - I_{R0}(e^{qV_{BC}/k_BT} - 1) \tag{3.138b}$$

关断阻抗 在截止状态,集电极电流为

$$I_C \simeq -\alpha_F I_{F0} + I_{R0} \tag{3.139}$$

利用式(3.79),得出

$$I_C = \frac{I_{CBO} - \alpha_F I_{EBO}}{1 - \alpha_F \alpha_R} \tag{3.140}$$

所以,关断阻抗近似为

$$R_{off} = \frac{V_{CE}}{I_C} \simeq \frac{V_{CE}(1 - \alpha_F \alpha_R)}{I_{CBO} - \alpha_F I_{EBO}} \tag{3.141}$$

导通阻抗 在饱和状态下,由于发射结和集电结都是正偏,降落在晶体管上的电压 V_{CE}很小,集电极电流 I_C 基本上由外电路决定而近似常量(V_{CC}/R_L),所以用它作为独立变量是方便的。这时,从式(3.138)可得

$$V_{BE} = \frac{k_BT}{q}\ln\left[\frac{I_B + (1 - \alpha_R)I_C}{I_{EBO}}\right] \tag{3.142a}$$

$$V_{BC} = \frac{k_BT}{q}\ln\left[\frac{\alpha_F I_B - (1 - \alpha_F)I_C}{I_{CBO}}\right] \tag{3.142b}$$

写出上面二式时用到了 $I_E + I_B + I_C = 0$ 的条件。由式(3.142)可得到导通阻抗为

$$R_{on} = \frac{V_{CE,饱和}}{I_C} = \frac{V_{BE} - V_{BC}}{I_C}$$

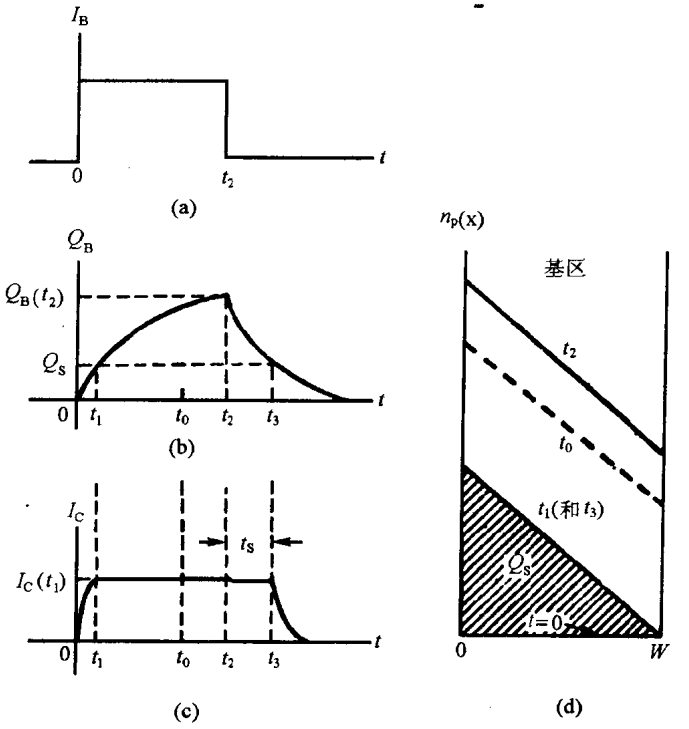
$$= \frac{k_BT}{qI_C}\ln\left\{\frac{1 + (1 - \alpha_R)\dfrac{I_C}{I_B}}{\alpha_R\left[1 - \left(\dfrac{1 - \alpha_F}{\alpha_F}\right)\dfrac{I_C}{I_B}\right]}\right\} \tag{3.143}$$

从式(3.141)可知,结的反向饱和电流 I_{CBO}和 I_{EBO}小时,关断阻抗就高。导通阻抗[式(3.143)]近似地反比于集电极电流 I_C,当 I_C 很大时,导通阻抗很小。实际上基区和集电区的欧姆电阻应包含在总阻抗内,特别对于导通阻抗更是这样。

3.5.3 导通时间和存储延迟时间

双极晶体管的开关时间是器件从关断状态到导通状态或从导通状态到关断状态所需的时间。当一个电流脉冲在时间 $t = 0$ 时加到基极-发射极端(见图 3.22(a)),晶体管被"接通"。从图 3.22 看到,晶体管作为开关应用时,输入信号 I_B 和输出信号 I_C 变化的幅度都很大,并且在截止和饱和之间近似于突变。对于这种大信号瞬变问题的分析通常采用电荷控制理论。在电荷控制理论中,受控制的不是电流或电压,而是利用各个区的受控电荷来建立方程,这些受控电荷包括发射区、基区和集电区内的存储电荷,以及发射结和集电结的耗尽区电荷。下面的讨论只考虑基区和集电区内的存储电荷。发射区内的存储电荷量很小,是可以忽略的。所以在我们的做法中,主要的问题是没有考虑结的充放电作用,或者说在开关时间中没有考虑导通过程中的延迟时间和关断过程中的下降时间。实际上,从基极输入脉冲到发射结接近导通(I_C 开始从零上升)需要一段时间,在此期间内必须对结充电,使发射结正偏;而从导通状态向关断

状态转换中，除了移走存储电荷需要所谓的存储延迟时间外，I_C 还必须经过一段时间才能下降到接近于零，在此期间内通过结电容放电，集电结从零偏变到反偏，这段时间称为下降时间。



3.22 (a) 基极输入电流脉冲；(b) 基区存储电荷随时间的变化；(c) 集电极电流随时间的变化；(d) 基区内少数载流子在不同时间的分布

1. 导通时间

在导通期间，基极电流是常数[图 3.22(a)]，其大小为

$$I_B = \frac{V_s - V_{BE}}{R_s} \simeq \frac{V_s}{R_s} \tag{3.144}$$

V_s 为输入脉冲幅度，R_s 是输入端的偏置电阻。基区存储电荷 Q_B 的变化由 I_B 和基区复合电流 Q_B/τ_n（τ_n 是基区电子的寿命）引起，电荷控制方程为

$$\frac{dQ_B}{dt} = I_B - \frac{Q_B}{\tau_n} \tag{3.145}$$

若在导通前 $Q_B=0$，则方程(3.145)的解为

$$Q_B(t) = I_B\tau_n(1 - e^{-t/\tau_n}) \tag{3.146}$$

$Q_B(t)$曲线表示在图3.22(b)中。集电极电流可以从 $Q_B=\tau_B I_C$ 得到：

$$I_C(t) = \frac{I_B\tau_n}{\tau_B}(1 - e^{-t/\tau_n}) \tag{3.147}$$

$I_C(t)$曲线画在图 3.22(c)中。

在输入脉冲有足够高度和宽度的情形下，晶体管将进入饱和状态。设在 $t = t_1$ 时刻达到

临界饱和($V_{BC}=0$),相应的集电极电流为

$$I_{CS}=\frac{V_{CC}-V_{CE}}{R_L}\simeq\frac{V_{CC}}{R_L} \tag{3.148}$$

临界饱和时基区存储的电荷量 Q_S 可表示为

$$Q_S=\tau_B I_{CS} \tag{3.149}$$

或

$$Q_S=I_B\tau_n(1-e^{-t_1/\tau_n}) \tag{3.150}$$

当 $Q_B>Q_S$ 时,晶体管工作在饱和状态,发射极电流和集电极电流都基本保持常数。图3.22(d)表明,$t>t_1$ 的任一时刻(比如说 $t=t_a$),电子 $n_p(x)$ 的分布将平行于 $t=t_1$ 时的分布,因此 $x=0$ 和 $x=W$ 处的浓度梯度不变,从而电流不变。由式(3.150)可以解出 t_1:

$$t_1=\tau_n\ln\left[\frac{1}{1-Q_S/I_B\tau_n}\right] \tag{3.151}$$

因此,为了减小导通时间 t_1,需缩短少数载流子(电子)寿命 τ_n,减小 Q_S 及增大 I_B。

达到临界饱和所需的基极电流为

$$I_{BS}=\frac{I_{CS}}{\beta_0} \tag{3.152}$$

I_B 通常超过 I_{BS},过量基极电流为

$$I_{BX}=I_B-I_{BS} \tag{3.153}$$

I_{BX}驱使晶体管进入深饱和状态,即发射结和集电结都正偏的状态。由于 $V_{BC}>0$,集电结也发生注入现象,使基区内有过量存储电荷 Q_B',集电区内有过量存储电荷 Q_C'。稳定时,过量基极电流将完全补偿这些过量存储电荷的复合损失,故有

$$I_{BX}=\frac{Q_B'}{\tau_n}+\frac{Q_C'}{\tau_{pc}} \tag{3.154}$$

τ_{pc}为集电区少数载流子(空穴)寿命。

2. *存储延迟时间*

在 t_2 时刻,基极电流突然下降到零,Q_B' 和 Q_C' 开始减少,设在 t 时刻它们下降到零。从 t_2 至 t_3 这段时间称为存储延迟时间,我们以 t_s 表示,在此期间有

$$-\frac{d}{dt}(Q_B'+Q_C')=\frac{Q_S+Q_B'}{\tau_n}+\frac{Q_C'}{\tau_{pc}} \tag{3.155}$$

若双极晶体管经过两次扩散杂质形成且集电区的掺杂浓度远低于基区(故 $Q_C'\gg Q_B'$),则上述方程简化为

$$-\frac{d}{dt}Q_C'=\frac{I_{CS}}{\beta_0}+\frac{Q_C'}{\tau_{pc}} \tag{3.156}$$

Q_C' 应满足下述边界条件:$t=t_2$ 时,$I_B-\frac{I_{CS}}{\beta_0}=\frac{Q_C'}{\tau_{pc}}$;$t=t_s$ 时,$Q_C'=0$。从 t_2 到 t_3 积分方程(3.156)以后得到

$$t_s=\tau_{pc}\ln\left[\frac{I_B}{I_{CS}/\beta_0}\right] \tag{3.157}$$

所以小的 τ_{pc}或 I_B 将使 t_s 下降。另外,器件进入饱和区的深度越小,存储时间也将缩短。从

t_3 开始，晶体管脱离饱和状态，I_C 随时间迅速下降到零，所以 t_s 实际上就是晶体管的关断时间。

综上所述，我们用电荷控制方程讨论了开关的瞬态过程。导通时间取决于如何迅速把电子(n-p-n 晶体管基区的少数载流子) 注入到基区，关断时间取决于如何通过复合使少数载流子(对于 n-p-n 晶体管并且集电区低掺杂的情形主要是集电区的空穴)迅速消失。开关晶体管最重要的参数之一是少数载流子的寿命。对于高速开关管，降低少数载流子寿命的一个有效方法是在基区和集电区引进复合中心(深能级杂质)。

3.6 异质结双极晶体管(HBT)

异质结是两种不同的半导体之间形成的结，例如在 p 型 GaAs 上形成 n 型 $Al_xGa_{1-x}As$。$Al_xGa_{1-x}As$ 是 AlAs 和 GaAs 这两种Ⅲ-Ⅴ族化合物半导体固溶形成的合金，x 是 AlAs 在合金中的摩尔分数。室温(300K)时，AlAs 的禁带宽度是 2.15eV，GaAs 的禁带宽度是 1.42eV，它们固溶形成的二元合金 $Al_xGa_{1-x}As$ 的禁带宽度比 GaAs 的大。异质结具有许多独有的性质，这些特性是前面讨论的常规半导体 p-n 结(同质结)所不具备的。异质结已经得到了许多重要应用，特别是在光电器件和量子效应器件方面。我们将在以后的有关章节中讨论这些器件。本节讨论异质结双极晶体管的基本工作原理。

3.6.1 异质结的能带图

图 3.23(a)表示形成异质结之前分离的两块半导体(n 型 $Al_xGa_{1-x}As$ 和 p 型 GaAs)的能带图。这两块半导体有不同的禁带宽度 E_g，不同的介电常数 ε_s，不同的功函数 $q\phi_s$ 及不同的亲和能 $q\chi$。功函数定义为将一个电子从费米能级 E_F 拿到材料外面(真空能级)所需要的能量，电子亲和能是把一个电子从导带底 E_C 移到真空能级所需的能量。我们对上述各量加下标以示区别，下标"1"代表窄禁带半导体，下标"2"代表宽禁带半导体。两种半导体导带边缘的能量差用 ΔE_C 表示；价带边缘的能量差用 ΔE_v 表示。如图 3.23(a)所示，$\Delta E_C = q(\chi_1 - \chi_2)$。

图 3.23(b)表示由这两种半导体形成理想突变异质结在热平衡状态下的能带图。图中我们假定两种异质半导体界面上的陷阱和产生-复合中心可以忽略。这个假设只有当两个半导体的晶体常数严格匹配或形成所谓应变层结构时才是正确的。最重要的异质结材料是Ⅲ-Ⅴ族化合物半导体，如 GaAs 及其三元化合物 $Al_xGa_{1-x}As$ 之类的固溶体，其中 x 可以从 0 变化到 1。当 $x=0$ 时为 GaAs，在 300K 下禁带宽度为 1.42eV，晶格常数为 5.6533Å；当 $x=1$ 时为 AlAs，它的禁带宽度为 2.15eV，晶格常数为 5.6605Å。三元化合物 $Al_xGa_{1-x}As$ 的禁带宽度随 x 的增加而增加，而晶格常数基本上保持不变，甚至在 $x=0$ 和 $x=1$ 的极端情况下，晶格常数的失配也仅有 0.1%。能带图的构成要满足下面两个基本要求：(1) 热平衡下界面两边的费米能级必须相同；(2) 真空能级必须是连续的，且平行于能带边缘。因此，只要禁带宽度 E_g 和亲和能 $q\chi$ 不是掺杂浓度的函数(即半导体是非简并的)，导带边缘的不连续性 ΔE_c 和价带边缘的不连续性 ΔE_v 也就不受掺杂的影响了。总的内建电势 V_{bi} 等于两部分内建电势之和 $(V_{b1} + V_{b2})$，V_{b1} 和 V_{b2} 分别是半导体 1 和 2 在热平衡时的静电势。

任意偏置下耗尽层宽度和电容均可解突变结界面每一边的泊松方程得到。一个边界条件是电位移的连续性，即 $\varepsilon_1\mathscr{E}_1 = \varepsilon_2\mathscr{E}_2$，$\mathscr{E}_1$ 和 $\mathscr{E}_2$ 分别是半导体 1 和 2 在界面处($x=0$)的电场强度。

我们得到

$$x_1 = \left[\frac{2N_D\varepsilon_1\varepsilon_2(V_{bi}-V)}{qN_A(\varepsilon_1N_A+\varepsilon_2N_D)}\right]^{1/2} \tag{3.158}$$

$$x_2 = \left[\frac{2N_A\varepsilon_1\varepsilon_2(V_{bi}-V)}{qN_D(\varepsilon_1N_A+\varepsilon_2N_D)}\right]^{1/2} \tag{3.159}$$

和

$$C = \left[\frac{qN_AN_D\varepsilon_1\varepsilon_2}{2(\varepsilon_1N_A+\varepsilon_2N_D)(V_{bi}-V)}\right]^{1/2} \tag{3.160}$$

式中 V 为加在结上的总电压,分配在两种半导体上的电压有下面的关系:

$$\frac{V_{b1}-V_1}{V_{b2}-V_2} = \frac{N_D\varepsilon_2}{N_A\varepsilon_1} \tag{3.161}$$

其中 $V=V_1+V_2$。

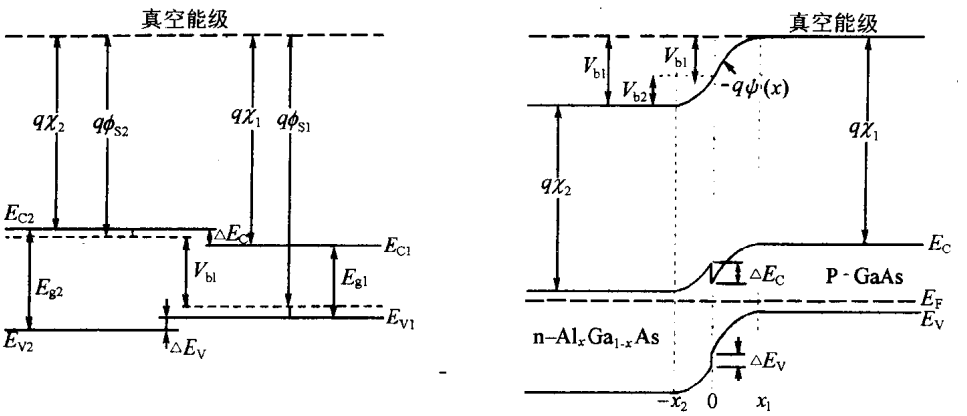


图 3.23 (a) 两块孤立半导体(n-$Al_xGa_{1-x}As$ 和 p-GaAs)的能带;(b) 热平衡下理想 n-p 异质结的能带图

下面着重讨论图 3.23(b)所示能带结构的基本特点。对于理想突变异质结,界面两侧的电势分布 $\psi(x)$满足下述泊松方程:

$$\frac{d^2\psi}{dx^2} = \begin{cases} -\dfrac{qN_D}{\varepsilon_2} & (-x_2 \leqslant x \leqslant 0) \\ \dfrac{qN_A}{\varepsilon_1} & (0 \leqslant x \leqslant x_1) \end{cases} \tag{3.162}$$

选择 $x=x_1$ 为电势零点,即 $\psi(x_1)=0$,且 $\psi(-x_2)=V_{bi}$。对方程(3.162)积分二次,并注意在耗尽区边界 $x=-x_2$ 和 $x=x_1$ 处电场是零(即 $d\psi/dx=0$)以及在界面 $x=0$ 处 $\psi(x)$连续,可得

$$\psi(x) = \begin{cases} V_{bi} - \dfrac{qN_D(x+x_2)^2}{2\varepsilon_2} & (-x_2 \leqslant x \leqslant 0) \\ \dfrac{qN_A(x+x_1)^2}{2\varepsilon_1} & (0 \leqslant x \leqslant x_1) \end{cases} \tag{3.163}$$

静电势能[$-q\psi(x)$]为抛物线，在图 3.23(b)中由真空能级随位置 x 的变化表示。不难看出，导带边 $E_c(x)$可以通过真空能级(即静电势)和亲和能 $q\chi(x)$表示如下：

$$E_c(x) = -q\psi(x) - q\chi(x) \tag{3.164}$$

对于我们考虑的突变异质结，因为 $\chi(x)$在界面($x=0$)处突变，该处形成一个势垒尖峰，它将阻碍电子通过界面。

上面的讨论可以推广到缓变异质结。从能带的角度，缓变指禁带宽度 E_g 和电子亲和能 $q\chi$ 是位置 x 的函数，也就是 ΔE_c 和 ΔE_v 都是 x 的函数。在几十纳米的距离上改变合金材料的组分比例将得到一个缓变异质结。例如 $Al_xGa_{1-x}As$，如果组分比例 x 从某一数值逐渐下降到零，则禁带将从较大的宽度缓变到 GaAs 的禁带宽度，电子亲和能则具有相反的变化趋势。对于 n-p-n 晶体管，我们特别关心电子的流动，也就是特别关心导带边 $E_c(x)$随距离 x 的变化。根据方程(3.164)，这不难从静电势能[$-q\psi(x)$]和电子亲和能 $q\chi$ 得到，如图 3.24 所示。在图中，$q\chi$ 是 $Al_xGa_{1-x}As$ 的亲和能。从该图看出，界面处的势垒尖峰基本消失。

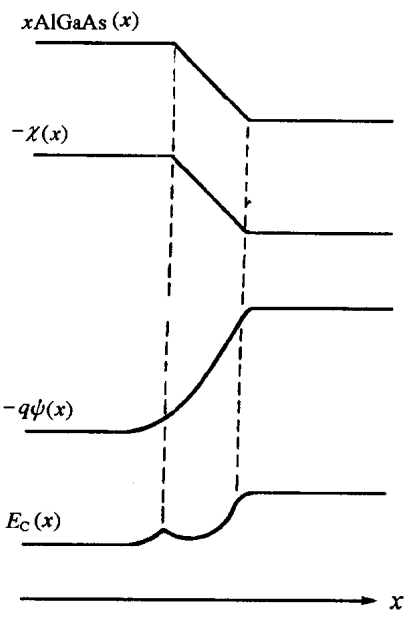


图 3.24　缓变异质结中导带边 $E_c(x)$随位置 x 的变化与组分、电子亲和势和静电势的关系

3.6.2　HBT 电流放大的基本理论

图 3.25(a)是具有宽禁带发射区的 HBT 的热平衡能带图。这种器件用 n 型 $Al_xGa_{1-x}As$ 作发射区，p 型 GaAs 作基区及 n 型 GaAs 作集电区。为了下面的讨论方便，假定发射结是缓变异质结。图 3.25(b)是该晶体管工作在放大状态下的能带图。从该图可以看出，与电子从发射区向基区注入相比，空穴由基区向发射区注入时要克服一个附加的台阶 ΔE_v，这就使发射结的注入比 I_{En}/I_{Ep}很高，即发射效率很高。显然，对于 n-p-n 晶体管，只有价带边能量差 ΔE_v 很大的异质结才有很高的注入比。

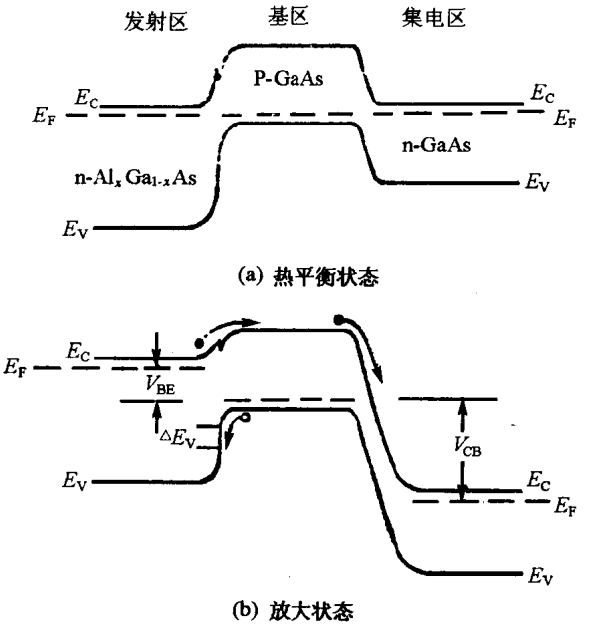


图 3.25　n-p-n 异质结双极型晶体管的能带图

在放大状态下，晶体管的基极电流主要来自以下几个方面：发射结耗尽层内的复合电流 I_{ER}，基区内复合而必须补偿空穴损失的电流 I_{BR}，基区向发射区注入的空穴电流 I_{EP}；集电极电流主要来自发射结注入并穿过基区的电流 I_{Cn}。于是共发射极电流增益可表示为

$$\beta_0 = \frac{I_{Cn}}{I_{Ep} + I_{BR} + I_{ER}} < \beta_{max} = \frac{I_{En}}{I_{Ep}} \tag{3.165}$$

这里 β_{max}表示注入比，也是晶体管受注入比限制时的最大电流增益。

根据晶体管理论，有

$$I_{En} = qA\frac{D_n n_{p0}}{W}(e^{qV_{BE}/k_BT} - 1) \tag{3.166}$$

$$I_{Ep} = qA\frac{D_{pE} p_{e0}}{L_{pE}}(e^{qV_{BE}/k_BT} - 1) \tag{3.167}$$

利用 $n_{po} = n_{iB}^2/N_B$ 和 $p_{eo} = n_{iE}^2/N_E$，我们得到

$$\beta_{max} = \frac{D_n}{D_{pE}} \cdot \frac{L_{PE}}{W} \cdot \frac{N_E}{N_B} \cdot \frac{n_{iB}^2}{n_{iE}^2} \tag{3.168}$$

其中，D_n 为电子在基区的扩散系数，D_{PE}和 L_{PE}分别为空穴在发射区的扩散系数和扩散长度，W 是基区宽度，N_E 和 N_B 分别为发射区和基区的掺杂浓度，n_{iE}和 n_{iB}分别为发射区和基区的本征载流子浓度。如果忽略不同半导体材料之间的有效态密度 N_c，N_v 的差别，有

$$\frac{n_{iB}^2}{n_{iE}^2} = e^{(E_{gE} - E_{gB})/k_BT} = e^{\Delta E_g/k_BT} \tag{3.169}$$

E_{gE}和 E_{gB}分别为发射区材料和基区材料的禁带宽度。将上述比值代入式(3.168)，有

$$\beta_{max} = \frac{D_n}{D_{pE}} \cdot \frac{L_{PE}}{W} \cdot \frac{N_E}{N_B} e^{\Delta E_g / k_B T} \tag{3.170}$$

与掺杂分布的同质结晶体管(BJT)相比,β_{max}之比为

$$\frac{\beta_{max}(HBT)}{\beta_{max}(BJT)} = e^{\Delta E_g / k_B T} \tag{3.171}$$

其中 $\Delta E_g = E_{gE} - E_{gB}$,可见宽禁带发射区异质结可以使双极型晶体管的电流增益大幅度提高。通常选取 ΔE_g 大于 250meV($10k_BT$),与同质结情形相比,β_{max}提高 10^4 倍。这样,基区可以高掺杂,其浓度可高达 $10^{20}cm^{-3}$。基区高掺杂将使器件性能大大改善,主要表现在下述几个方面:第一,基区不容易穿通,从而厚度可以做得很小,即它不限制器件尺寸缩小;第二,基区电阻可以显著降低,从而最高振荡频率 f_{max}增加;第三,基区电导调制不明显,从而大电流密度时电流增益不会明显下降;第四,基区电荷对输出电压(集电结电压)不敏感,从而 Early 电压显著增大。另一方面,发射区的掺杂浓度可以很低($\sim 10^{17}cm^{-3}$),从而发射结耗尽层电容大大减小,器件的电流放大截止频率 f_T 提高。

利用外延技术可以制成各种各样的异质结双极晶体管。其中包括采用组分缓变材料作基区的器件(例如用 $Al_xGa_{1-x}As$ 作基区,x 值从发射结到集电结递减),以提供一个内建电场,减小基区渡越时间;也有采用双异质结构(宽禁带发射区和宽禁带集电区)的器件,以使发射结和集电结对称,改善放大状态和反转状态下的电流增益。

3.6.3 几类常见的 HBT

有多种材料体系可以制作异质结器件,下面介绍几类常见的异质结双极晶体管。

AlGaAs/GaAs HBT 这类 HBT 的发射区采用 $Al_xGa_{1-x}As$ 材料,Al(或 AlAs)的摩尔分数 x 选择在 0.25 左右,高于此值时 n 型 AlGaAs 中开始出现深施主,使发射结电容增加。$x = 0.25$时,发射区的禁带宽度比基区的大 0.30eV,注入效率可以显著提高。基区采用 p^+-GaAs 材料,典型厚度为 0.05~0.1μm,典型掺杂浓度 N_A 为 $5\times10^{18}\sim1\times10^{20}cm^{-3}$。集电区通常也采用 GaAs 材料($n^-$型)。集电区往下依次为 n^+-GaAs 埋层和半绝缘 GaAs(SI GaAs)衬底。SI GaAs 的获得是通过向 GaAs 引入深能级杂质(因而费米能级被钉扎在禁带中央)。这类 HBT 的一个重要优点是 $Al_xGa_{1-x}As$/GaAs 材料体系可以有良好的晶格匹配。由于 AlAs 和 GaAs 的晶格常数十分接近,而且它们热膨胀系数之间的差别也很小,无论怎样选择 Al 的摩尔分数 x 之值都能实现晶格匹配。其次,由于采用半绝缘衬底,器件之间容易隔离和互连,器件或互连线同衬底之间的电容可以忽略。特别在微波电路中,为了限制微波信号的衰减和降低电阻,互连线宽度较大,从而衬底的影响十分重要。单片微波集成电路用 GaAs 材料容易实现,而用 Si 材料则很难。此外,AlGaAs/GaAs 已被用来制作激光器、发光二极管、光探测器等光电子器件,这些器件同以 AlGaAs/GaAs HBT 为基础的电路可以单片集成。

InGaAs HBT 同 InP 晶格匹配的Ⅲ-Ⅴ族化合物半导体中包括 $In_{0.53}Ga_{0.47}As$(以下简写为 InGaAs)和 $In_{0.52}Al_{0.48}As$(简写为 InAlAs),InGaAs 的禁带宽度是 0.75eV,InAlAs 的禁带宽度是 1.5eV,而 InP 的禁带宽度是 1.35eV。用 InGaAs 作为基区而 InP 或 InAlAs 作为发射区构成 HBT,其主要优点是 InGaAs 中的电子迁移率很高,对于本征材料(无杂质散射),其电子迁移率是 GaAs 的 1.6 倍,Si 的 9 倍。这类器件的半绝缘衬底采用掺 Fe 的 InP。

$Si/Si_{1-x}Ge_x$ HBT 加入 Ge 会降低 Si 的禁带宽度,形成可以用于 HBT 基区的合金。由于

Ge 和 Si 的晶格常数(分别为 5.6575Å 和 5.4310Å)相差超过 4%，SiGe 合金的晶格常数将和 Si 的相差甚大，不可能实现晶格匹配。但是，如果合金层的厚度低于临界值，SiGe 合金和 Si 之间可以弹性调节，而不出现晶格失配，这就是所谓的应变层结构。实验表明，SiGe 合金层的厚度(基区宽度)超过 0.2μm 时，基极电流增加，这是失配位错所致。SiGe 合金和 Si 之间的禁带宽度差基本上是产生价带台阶 ΔE_v，这种情形对制作 n-p-n HBT 是十分有利的。SiGe 合金中 Ge 的摩尔分数已达到 20%，相应的禁带宽度差约为 200meV($8k_BT$)，由前面的讨论可知，这类 HBT 可以有很高的注入效率。近年来，随着分子束外延(MBE)和金属有机物化学汽相淀积(MOCVD)技术的发展，能够制作高质量应变层结构的 SiGe/Si 异质结，包括 HBT 在内的 SiGe/Si 异质结构器件受到广泛重视。与化合物异质结器件相比，这类器件由于采用成熟的硅工艺，工艺简单、可靠，价格便宜，机械和导热性能良好，并且可以在同一衬底上集成电子器件和光电子器件。

3.7 多晶硅发射极晶体管(PET)

尽管双极型晶体管具有截止频率高和跨导大等优越的电学特性，但长期以来双极芯片却存在功耗-延时乘积大和集成度低的缺点。近十多年来，由于采用了多晶硅发射极接触、自对准基极接触和深槽隔离等关键技术，高速和高密度的双极电路得到了迅速发展。图 3.26 表示一种先进的双极晶体管结构，自对准基极接触(这里采用 p^+-多晶硅)和深槽隔离大大减小了器件面积和寄生电容，从而显著地减小了双极电路的功耗-延时乘积和提高了集成度；多晶硅

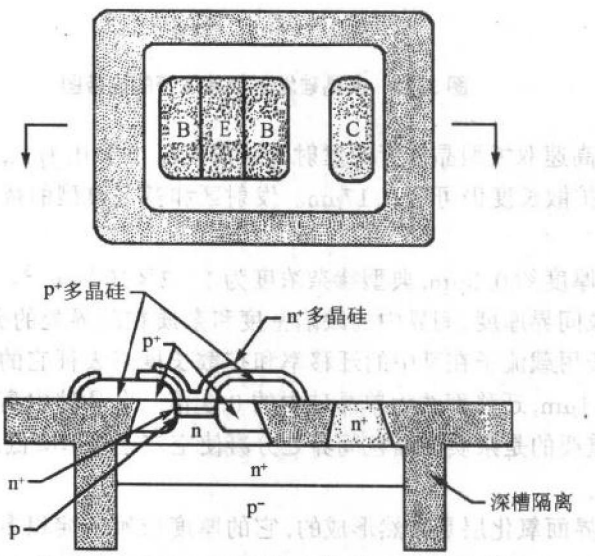


图 3.26 用双层多晶硅自对准和深槽隔离技术制作的硅双极晶体管的剖面图

发射极接触通过多晶硅/单晶硅界面或多晶硅层对少数载流子(在n-p-n晶体管的发射区中为空穴)的阻碍作用，使注入发射区的电流(基极电流)下降，电流增益大幅度提高，而电流增益的提高使得双极晶体管在不降低基区穿通电压和不损失电流增益的情况下可以通过提高基区掺杂浓度来减小基区宽度，实现器件的纵向按比例缩小。此外，如果晶体管的发射区是多晶硅中

的掺杂原子通过高温扩散向基区推进得到，则它同时实现了发射极自对准接触。自对准基极接触技术、深槽隔离技术以及多晶硅发射极接触的物理与工艺是近年来双极领域中最广泛的研究课题，本节只讨论与多晶硅发射极接触有关的物理问题和多晶硅发射极晶体管的电流放大原理。

3.7.1 能带图和物理参数

图 3.27 是多晶硅发射极晶体管的能带示意图。不同于常规的双极型晶体管，其发射区包括晶态硅(c-Si)和多晶硅(poly-Si)两个区域，二者之间通常有一个厚度 δ 十分小的界面氧化层。

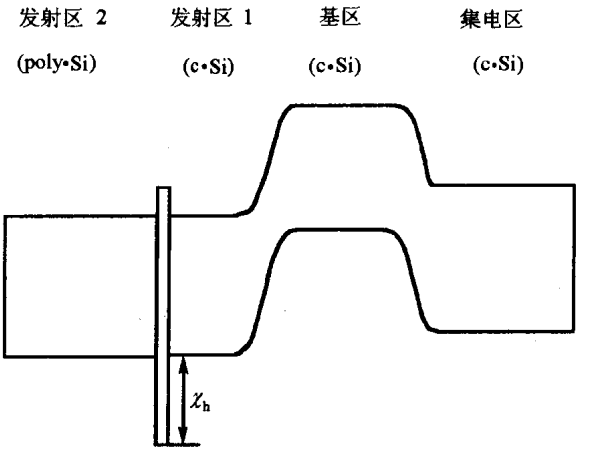


图 3.27 多晶硅发射极晶体管的能带图

晶态硅 目前高速双极型晶体管的发射区结深度已小于 0.1μm，而即使在掺杂浓度高至 $10^{20}cm^{-3}$时空穴的扩散长度仍可达 0.17μm。发射区和基区典型的掺杂浓度分别为 $10^{20}cm^{-3}$ 和 $10^{18}cm^{-3}$。

多晶硅 典型厚度约 0.3μm，典型掺杂浓度为 $1-2\times10^{20}cm^{-3}$。表征多晶硅性质的参数包括晶粒大小、晶粒间界厚度、间界中的缺陷密度和杂质在间界处的分凝度等，目前尚缺少具体的实验数据，一般用载流子在其中的迁移率和扩散长度来表征它的性质。空穴在多晶硅中的扩散长度约为 0.1μm，迁移率约为单晶硅中的 0.3 倍。多晶硅中高的掺杂浓度不仅减少器件的串联电阻，更重要的是杂质在晶粒间界处分凝使空穴迁移率降低，有利于减小基极电流和提高电流增益。

界面氧化层 界面氧化层是自然形成的，它的厚度依赖于淀积多晶硅前的清洗工艺。淀积前，硅片没有经过 HF 浸泡的(RCA 处理)，界面氧化层厚度约为 15Å；经过 HF 浸泡的(HF 处理)，界面氧化层往往残破不全，厚度约为 5Å。界面氧化层并不是理想的 SiO_2，而是 SiO_x(x 约为 1.7)。对于完整的界面氧化层薄膜，载流子以隧道效应方式通过，但薄层对空穴构成的有效势垒高度比对电子构成的大得多，图 3.27 说明了这一点。界面氧化薄层势垒的高度与厚度有关，对于空穴的势垒，氧化层厚度 25Å 时势垒高度约为 1eV，厚度 15Å 约为 0.5eV，它们远小于热生长氧化层形成的势垒高度(约 3eV)。

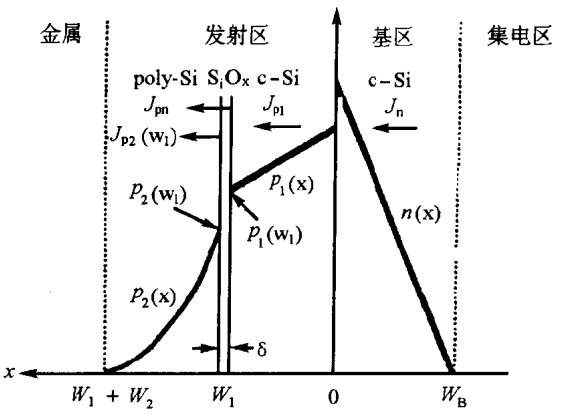


图 3.28　PET 中非平衡少子分布和电流

3.7.2　少子分布和电流密度

多晶硅发射极晶体管各个区域中非平衡少子分布及电流密度示于图 3.28 中。

c-Si 发射区　忽略复合，则电流密度 J_{P1} 为常数，由下式表示：

$$J_{P1} = -qD_{p1}\frac{P_1(W_1) - p_1(0)}{W_1} \tag{3.172}$$

其中，

$$P_1(0) = \frac{n_i^2}{N_D}e^{qV_{BE}/k_BT} \tag{3.173}$$

N_D 是 c-Si 发射区的掺杂浓度，假定为常量；V_{BE} 是加在发射结上的电压；$P_1(W_1)$ 是界面层（$x = W_1$）处 c-Si 一侧的空穴浓度，待确定。

poly-Si 发射区　这一区域的厚度 W_2 约等于 0.3μm，空穴在其中的扩散长度 L_{P2} 近似等于 0.1μm，所以必须考虑载流子的复合。假定杂质分布均匀，则空穴连续方程为

$$\frac{d^2P_2(x)}{dx^2} - \frac{P_2(x)}{L_{P2}^2} = 0 \tag{3.174}$$

金属接触处 $P_2(W_1 + W_2) = 0$，所以上述方程的解为

$$P_2(x) = \frac{P_2(W_1)}{\sinh(W_2/L_{P2})}\sinh\left(\frac{W_1 + W_2 - x}{L_{P2}}\right) \tag{3.175}$$

界面处的空穴电流密度为

$$J_{P2}(W_1) = \frac{qD_{P2}P_2(W_1)}{L_{P2}}\coth\left(\frac{W_2}{L_{P2}}\right) \tag{3.176}$$

$P_2(W_1)$ 是界面（$x = W_1$）处 poly-Si 一侧的空穴浓度，待确定。

界面区　为简化分析，忽略界面层内的复合和界面层上的电压，于是空穴只以隧穿方式通过界面，隧穿几率可表示为

$$T_P(E_x) = e^{-\beta\delta\sqrt{\chi_h - E_x}} \tag{3.177}$$

其中 $\beta = 4\pi\sqrt{2m_p}/h$，$\delta$ 是界面层厚度，χ_h 是界面对于空穴的势垒高度，E_x 是空穴垂直于界

面运动的能量，χ_h 和 E_x 都以价带顶作为零点。只是 E_x 接近价带顶的空穴对电流有显著贡献，故可以把 $T_P(E_x)$ 的指数项在 $E_x=0$ 附近作级数展开，并只保留常数项和线性项，于是有

$$T_P(E_x) = e^{-b_h + C_h E_x} \tag{3.178}$$

其中

$$b_h = \frac{4\pi\delta}{h}\sqrt{2m_p\chi_h}, \quad C_h = \frac{2\pi\delta}{h}\sqrt{\frac{2m_p}{\chi_h}} \tag{3.179}$$

定义隧穿界面的粒子数与到达界面的粒子数之比为平均隧穿几率，并以 α 表示，则在粒子按能量为玻耳兹曼分布(玻耳兹曼因子 e^{-E_x/k_BT})的条件下，可求得

$$\alpha = \frac{e^{-b_h}}{1 - C_h k_B T} \tag{3.180}$$

基区 实验表明，薄界面层对电子构成的势垒较低以致隧穿几率很大，加之发射区重掺杂使电子浓度很大，故电子电流基本上不受界面层的影响。注入基区的电子电流与常规情形一样，其密度可由下式表示：

$$J_n = \frac{qD_n n_i^2 e^{qV_{BE}/k_BT}}{W_B N_A} \tag{3.181}$$

N_A 为基区掺杂浓度。

3.7.3 注入发射区的空穴电流密度 J_{P1}

J_{P1} 由(3.172)式表示，为了确定界面处的空穴浓度 $P_1(W_1)$，应对有、无多晶硅层二种情形分别考虑，前者对应 $W_2 \neq 0$，W_2 是多晶硅层厚度。

(1) $W_2 \neq 0$

这时，空穴隧穿界面层受到在多晶硅层中扩散过程限制，$x = W_1$ 处的边界条件可表示为

$$P_2(W_1) = \alpha P_1(W_1), \quad J_{P2}(W_1) = J_{P1}(W_1) \tag{3.182}$$

空穴电流在界面处连续是因为忽略了界面层内复合的结果。把上述边界条件用于 $J_{P2}(W_1)$ 的表示式(3.176)，解出 $P_1(W_1)$ 后代入 J_{P1} 的表示式(3.172)，即得

$$J_{P1} = \frac{qD_{P1} n_i^2 e^{qV_{BE}/k_BT}}{W_1 N_D}\left[1 + \frac{1}{\alpha}\cdot\frac{D_{P1}L_{P2}}{D_{P2}W_1}\tanh\left(\frac{W_2}{L_{P2}}\right)\right]^{-1} \tag{3.183}$$

(2) $W_2 = 0$(*MIS* 结构)

这时，电流受空穴隧穿界面层的速度限制，故有

$$J_{P1} = qv_c\alpha P_1(W_1) \tag{3.184}$$

式中 $v_c = \sqrt{k_BT/2\pi m_p}$ 是空穴的热发射速度，由上式解出 $P_1(W_1)$ 后代入式(3.172)，得到

$$J_{P1} = \frac{qD_{P1} n_i^2 e^{qV_{BE}/k_BT}}{W_1 N_D}\left[1 + \frac{D_{P1}}{v_c W_1}\cdot\frac{1}{\alpha}\right]^{-1} \tag{3.185}$$

3.7.4 电流增益

在现代先进的工艺水平下，集电极电流基本上等于注入基区的电流，基极电流基本上等于注入发射区的电流。对于 n-p-n 晶体管，电流增益可以表示成

$$\beta = J_n/J_p \tag{3.186}$$

对于多晶硅发射极晶体管，我们引入归一化的电流增益 γ 来表示电流增益依赖于发射区结构的趋势。γ 定义为

$$\gamma \equiv \beta/\beta(W_2 = 0) \tag{3.187}$$

$\beta(W_2=0)$是假设晶体管没有多晶硅发射区时的电流增益，显然 γ 又可表示成

$$\gamma = J_{P1}(W_2 = 0)/J_{P1} \tag{3.188}$$

将式(3.185)代入上式，有

$$\gamma = 1 + \frac{1}{\alpha} \cdot \frac{D_{P1}}{D_{P2}} \frac{L_{P2}}{W_1} \tanh\left(\frac{W_2}{L_{P2}}\right) \tag{3.189}$$

这一结果说明，由于薄界面氧化层($\alpha<1$)和多晶硅层($D_{P2}<D_{P1}$)的存在，多晶硅发射极晶体管的电流增益比常规晶体管得到改进，*RCA* 器件应当属于这种情形。

在没有界面氧化层($\alpha=1$)的情形下，当 $W_2>L_{P2}[\tanh(W_2/L_{P2})\simeq 1]$时，式(3.189)简化为

$$\gamma \simeq 1 + \frac{D_{P1}}{D_{P2}} \cdot \frac{L_{P2}}{W_1} \tag{3.190}$$

电流增益改进是由于空穴在多晶硅中的迁移率低($D_{P2}<D_{P1}$)，根据前述，γ 的典型数值近似为 4。HF 器件大致属于这种情形。

3.8 p-n-p-n 结构

这种结构可用作二端、三端或四端器件。作为二端器件，它具有开关性质，如果在内部 p 区加第三个电极，就构成三端器件，开关过程由通过第三个电极的电流控制。三端的 p-n-p-n 器件称为半导体可控整流器，或可控硅，因为它实际上只用硅材料制作。具有 p-n-p-n 结构的器件统称晶闸管(Thysistor)。依赖于器件结构，晶闸管的开态电流可高达几百至几千安培，关态电压可达几千甚至上万伏。就像集成电路是微电子技术的基础一样，晶闸管是电力电子技术的重要基础。本节只讨论 p-n-p-n 二极管的基本特性。

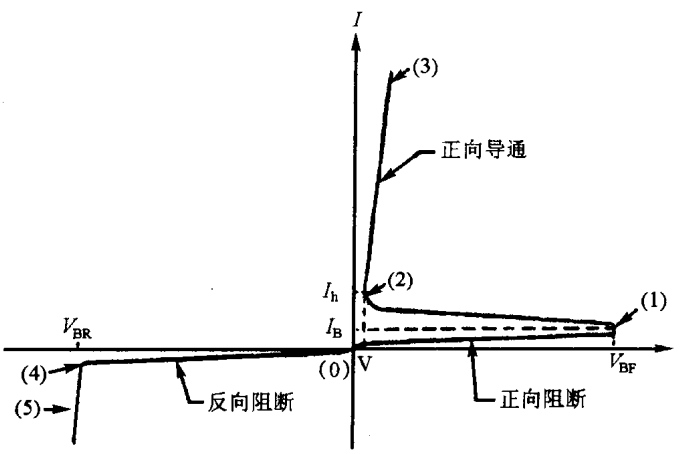


图 3.29 p-n-p-n 二极管的电流-电压特性

p-n-p-n 二极管基本的电流-电压特性如图 3.29 所示，它有 5 个不同的区域：

(0)—(1)　正向高阻区

(1)—(2)　负阻区

(2)—(3)　正向低阻区

(0)—(4)　反向关断区

(4)—(5)　反向击穿区

这样，在正向区工作的 p-n-p-n 二极管是一个双稳态器件，它能在高阻的关断态和低阻的导通态之间互相转换。

为了便于理解 p-n-p-n 二极管的正向特性，可以把它看成是两个背对背连接的晶体管，一个是 p-n-p 管(BJT_1)，另一个是 n-p-n 管(BJT_2)，中间的 p 区和 n 区为两个晶体管共有，如图 3.30所示。这样，BJT_1 的基区和 BJT_2 的集电区相连，即 $I_{B1} = I_{C2}$；而 BJT_1 的集电区和 BJT_2 的基区相连，即 $I_{C1} = I_{B2}$。

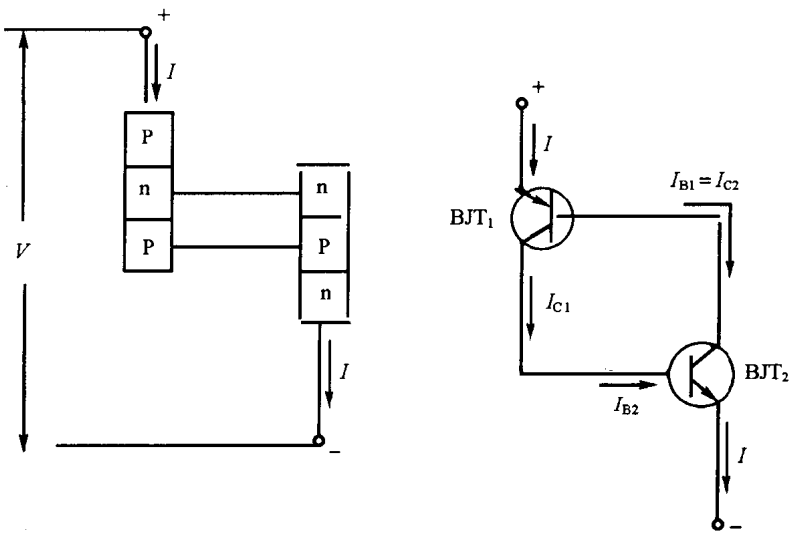


图 3.30　p-n-p-n 二极管的等效电路

令两个晶体管的电流增益分别为 α_1 和 α_2，漏电流分别为 I_{CBO1}和 I_{CBO2}，则有

$$I_{B1} = (1 - \alpha_1)I_{E1} - I_{CBO1}$$

$$I_{C2} = \alpha_2 I_{E2} + I_{CBO2}$$

利用 $I_{B1} = I_{C2}$，并注意 $I_{E1} = I_{E2} = I$，得到

$$I = \frac{I_{CBO}}{1 - (\alpha_1 + \alpha_2)}$$

其中 $I_{CBO} = I_{CBO1} + I_{CBO2}$。小电流时 α_1 和 α_2 都比 1 小得多，流过器件的电流就是漏电流 I_{CBO}。当外加电压增加到正向转折电压 V_{BF}时，$\alpha_1 + \alpha_2$ 接近于 1，流过器件的电流迅速增大。

正向关断时，p-n-p-n 二极管外侧的两个 p-n 结 J_1 和 J_3 正向偏置，中间的 p-n 结 J_2 反向偏置，外加电压主要降落在 J_2 上，如图 3.31 所示。发生正向转折以后，由于 $\alpha_1 + \alpha_2 > 1$，从 J_1 和 J_3 注入内部 p 区和 n 区的载流子总数$(\alpha_1 + \alpha_2)I/q$ 中只有 I/q 能通过J_2 结，结果 J_2 结因

为两侧载流子堆积而变成正偏，因此器件上的电压突然下降，器件进入正向导通状态。这时，三个结都处在正向偏置，两个晶体管都处在饱和状态，整个器件上的电压很低，等于（$V_1 - |V_2| + V_3$），它近似等于一个 p-n 结的导通电压。

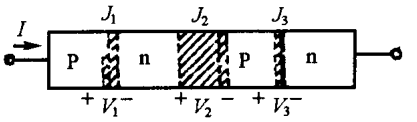


图 3.31　正向关断时 p-n-p-n 二极管的电压降和耗尽区

在反向关断状态，J_2 结正偏，J_1 结和 J_3 结反偏，内部 n 区一般是高阻区，所以反偏时外加偏压主要降落在 J_1 结上，如图 3.32 所示。

如果在内部 p 区上制作第三个电极作为控制极，就构成了可控整流器件。通过控制极加一个触发信号时，等效的 n-p-n 管首先导通，信号电流被放大，随即驱动等效的 p-n-p 管导通。由于两个晶体管的集电极电流分别流向另一晶体管的基极，所以器件被信号触发以后就维持导通，而不需要触发信号继续存在，只有加在两端的电压下降到零或反向时器件才关断。

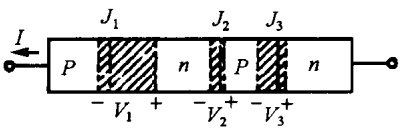


图 3.32　反向关断时 p-n-p-n 二极管的耗尽区和压降

寄生的 p-n-p-n 结构可能给半导体器件带来很大麻烦，最典型的例子是 CMOS 反相器中出现的闩锁(latch-up)效应。所谓闩锁效应，是寄生于 CMOS 器件中的 p-n-p-n 结构在某些不稳定因素的触发下正向导通，有很大的电流通过，引起器件失效。

第四章　化合物半导体场效应晶体管

本章主要考虑化合物半导体场效应晶体管。场效应晶体管(FET)是单极器件,即实际上只有多数载流子参加导电的器件。

4.1　肖特基势垒和欧姆接触

4.1.1　肖特基势垒

金属和半导体接触时,由于金属的功函数一般和半导体的功函数不同而存在接触电势差,结果在接触界面附近形成势垒,通常称为肖特基势垒。下面以金属和 n 型半导体接触为例,来讨论肖特基势垒。

1. 理想情形

图 4.1 是理想的金属-半导体接触的能带图。所谓理想情形,指接触处的半导体表面不存

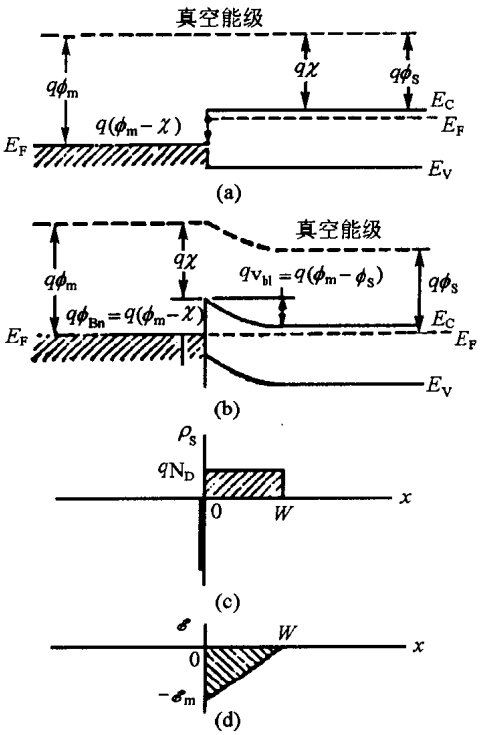


图 4.1　(a) 邻近的孤立金属和孤立 n 型半导体在非平衡条件下的能带图;
(b) 金属-半导体接触在热平衡下的能带图;(c) 电荷分布;(d) 电场分布

在表面态。图中 $q\phi_m$ 和 $q\phi_s$ 分别为金属和半导体的功函数,$q\chi$ 为半导体的电子亲和能。功函数是把一个电子从费米能级移到材料外面(真空能级)所需要的能量,电子亲和能是把一个电

子从导带底移到真空能级所需要的能量。金属和半导体接触的势垒高度指电子从金属进入半导体必须克服的势垒的高度。对于金属和 n 型半导体的理想接触，势垒高度为

$$q\phi_{Bn} = q\phi_m - q\chi \tag{4.1}$$

$q\phi_{Bn}$(单位是 eV)或 ϕ_{Bn}(单位是 V)都表示势垒高度。n 型半导体的自建电压为

$$V_{bi} = \phi_m - \phi_s \tag{4.2}$$

ϕ_{Bn}和 V_{bi}之间的关系为

$$\phi_{Bn} = V_{bi} + V_n \tag{4.3}$$

qV_n 是半导体的导带底和费米能级之差。

对于金属和 n 型半导体的接触，金属一侧有负的表面电荷，半导体一侧存在等量的正空间电荷，这种电荷分布和具有同样电场分布的单边突变 p^+-n 结完全相同。于是我们得到半导体表面耗尽层的宽度为

$$W = \sqrt{\frac{2\varepsilon_s}{qN_D}(V_{bi} - V)} \tag{4.4}$$

正向偏置(即金属相对 n 型半导体加正电压)时，上式中的外加电压 V 取正值；反向偏置时，V 取负值。半导体内单位面积的空间电荷 Q_{sc}(C/cm^2)和单位面积耗尽层电容(F/cm^2)由下式给出：

$$Q_{sc} = qN_D W = \sqrt{2q\varepsilon_s N_D(V_{bi} - V)} \tag{4.5}$$

$$C = \left|\frac{\partial Q_{sc}}{\partial V}\right| = \sqrt{\frac{q\varepsilon_s N_D}{2(V_{bi} - V)}} = \frac{\varepsilon_s}{W} \tag{4.6}$$

对于金属和 p 型半导体的接触，可以进行与上述完全类似的讨论。金属和 p 型半导体理想接触的势垒高度为

$$q\phi_{Bp} = E_g - q(\phi_m - \chi) \tag{4.7}$$

E_g 是半导体的禁带宽度。因此，对于一种给定的半导体，任何金属在 n 型衬底和在 p 型衬底上的势垒高度之和总等于禁带宽度。

2. *界面态和势垒高度的一般表示式*

在实际的金属-半导体接触中，由于晶格不连续，在接触界面处产生大量的能量状态。这些能量状态叫做界面态或表面态，它们连续分布在禁带内。为了描述半导体的表面状态，引进中性能级 $q\phi_0$：当 $q\phi_0$ 以下的表面态全部被电子填充而以上的全部空出时，半导体表面是中性的。也就是说，低于 $q\phi_0$ 的界面态没有电子占据时带正电，其作用同施主类似；高于 $q\phi_0$ 的界面态被电子占据时带负电，其作用同受主类似。如果 $q\phi_0$ 和半导体的费米能级 E_F 重合，则界面态和半导体内部没有电子交换，界面的净电荷为零；如果 $q\phi_0 > E_F$，则电子从表面向体内转移，界面的净电荷是正的；如果 $q\phi_0 < E_F$，则电子从体内向表面转移，界面的净电荷是负的。根据以上分析，并考虑到在接触界面处几乎总有一层极薄的界面层 δ，可以得到如图 4.2 所示的实际金属-半导体接触的能带图。

设界面态的密度为 D_S(个/cm^2·eV)且与能量无关，则对于有受主型表面态($q\phi_0 < E_F$)的半导体，界面态电荷面密度(C/cm^2)为

$$Q_{ss} = -qD_s(E_g - q\phi_{Bn} - q\phi_0) \tag{4.8}$$

括号内的量简单地为表面处的费米能级与 $q\phi_0$ 之差。

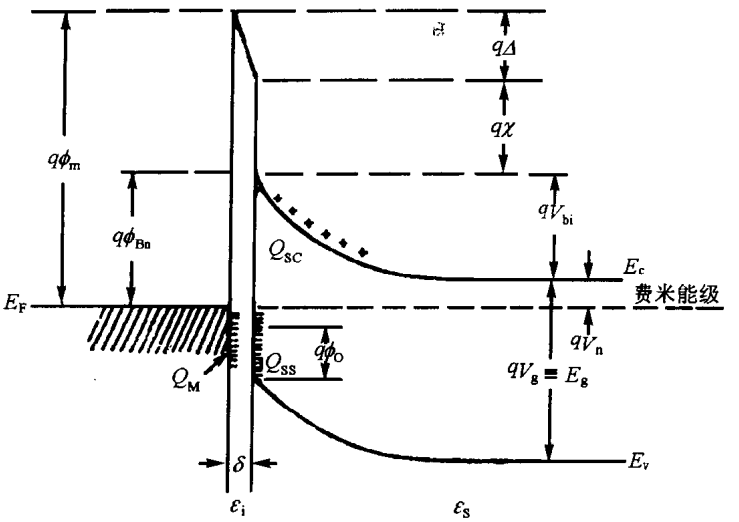


图 4.2　实际的金属-半导体接触

金属-半导体接触和单边突变 p-n 结的情形类似，半导体表面耗尽层电荷面密度为

$$Q_{sc} = \sqrt{2q\epsilon_s N_D V_{bi}} = \sqrt{2q\epsilon_s N_D(\phi_{Bn} - V_n)} \tag{4.9}$$

当界面层内不存在任何空间电荷效应(对于薄的界面层，这些效应可以忽略)时，根据电荷平衡条件，金属表面的电荷密度为

$$Q_M = -(Q_{ss} + Q_{sc}) \tag{4.10}$$

根据高斯定理，界面层内的电场强度为 Q_M/ϵ_i，ϵ_i 是界面层的介电常数。所以，界面层上的电势降落是

$$\Delta = -\delta \frac{Q_M}{\epsilon_i} \tag{4.11}$$

δ 是界面层厚度。

另一方面，从图 4.2 可得

$$\Delta = \phi_m - (\phi_{Bn} + \chi) \tag{4.12}$$

从方程(4.11)和(4.12)消去 Δ，并以方程(4.10)代替 Q_M，得到势垒高度的一般表示式为

$$\phi_{Bn} = c_1(\phi_m - \chi) + (1 - c_1)(E_g/q - \phi_0) - c_1 c_2(\epsilon_s/\epsilon_i)\delta \tag{4.13}$$

$$c_1 = \frac{\epsilon_i}{\epsilon_i + q^2 D_s \delta}, \quad c_2 = \sqrt{\frac{2qN_D V_{bi}}{\epsilon_s}} \tag{4.14}$$

当 $D_S \to 0$ 时，$C_1 \to 1$，而且因为 δ 上的电压降落通常可以忽略，有

$$q\phi_{Bn} \simeq q(\phi_m - \chi) \tag{4.15}$$

与理想情形一致。$D_s \to \infty$ 时，$c_1 \to 0$，有

$$q\phi_{Bn} \simeq E_g - q\phi_0 \tag{4.16}$$

势垒高度与金属的功函数无关。

4.1.2　肖特基势垒二极管

在金属-半导体接触中，电流输运主要由多数载流子承担。这与 p-n 结相反，p-n 结中的电

流输运主要由少数载流子承担。多数载流子通过势垒的方式包括扩散、热离子发射和隧道穿通，有赖于势垒区的有效宽度 d_T 和载流子的平均自由程 λ 之比。λ 可表示为电子热运动速度 $v_{th}=\sqrt{3k_BT/m_n}$ 和平均自由时间 $\tau=\frac{m_n\mu_n}{q}$ 的乘积，即

$$\lambda = \mu_n \sqrt{3k_BTm_n}/q \tag{4.17}$$

d_T 可考虑为从界面(势垒顶)算起势垒高度下降 k_BT(电子浓度增至 e 倍)的距离，设其上的电场强度近似等于界面处的场强 $\mathscr{E}_m=\sqrt{2qN_DV_{bi}/\varepsilon_s}$，则由 $q\mathscr{E}_m d_T=k_BT$，可得到

$$d_T \simeq \frac{1}{q}\sqrt{\frac{\varepsilon_s(k_BT)^2}{2qN_DV_{bi}}} \tag{4.18}$$

对于 $\lambda>d_T$，有

$$N_D > \varepsilon_s k_BT/6\mu_n^2 q m_n V_{bi} \tag{4.19}$$

式中，m_n 为半导体中电子有效质量，μ_n 是电子迁移率。对于 μ_n 大的半导体，例如 Si 和 GaAs，室温下掺杂浓度 $N_D>10^{12}cm^{-3}$ 时，上述不等式就能成立；但对迁移率小的半导体，例如非晶硅($\mu_n<10cm^2/V\cdot s$)，则必须有较高的掺杂浓度。

$\lambda>d_T$ 时，电子主要以热离子发射的方式通过势垒。所以在通常的金属-半导体接触中，多数载流子以热离子发射方式输运。

热离子发射电流密度可表示为

$$J = qn_sv_c \tag{4.20}$$

$v_c=\sqrt{k_BT/2\pi m_n}$(等于 $\bar{v}/4$，$\bar{v}=\sqrt{8k_BT/\pi m_n}$ 是热运动平均速率)为热发射速度，n_s 是界面处的电子浓度。n_s 与中性区(势垒底部)电子浓度 n_0 的关系为

$$\begin{aligned} n_s &= n_0 e^{-q(V_{bi}-V)/k_BT} \\ &= N_c e^{-q(\phi_{Bn}-V)/k_BT} \end{aligned} \tag{4.21}$$

V 是外加电压。上式的最后一步利用了 $n_0=N_c\times e^{-qV_n/k_BT}$ 和 $V_{bi}=\phi_{Bn}-V_n$。

这样，在外加电压 V 的作用下，从半导体流向金属的电流密度为

$$J_{S\to M} = qv_cN_ce^{-q(\phi_{Bn}-V)/k_BT} \tag{4.22}$$

而电子从金属向半导体运动的势垒高度总是 ϕ_{Bn}，与 V 无关，故有

$$J_{M\to S} = qv_cN_ce^{-q\phi_{Bn}/k_BT} \tag{4.23}$$

所以，在 V 作用下从半导体流向金属的净电流为

$$\begin{aligned} J &= J_{S\to M} - J_{M\to S} \\ &= J_S[e^{qV/k_BT}-1] \end{aligned} \tag{4.24}$$

其中

$$\begin{aligned} J_S &= qv_cN_ce^{-q\phi_{Bn}/k_BT} \\ &= q\left(\frac{k_BT}{2\pi m_n}\right)^{1/2}\cdot 2\left(\frac{2\pi m_n k_BT}{h^2}\right)^{3/2} e^{-q\phi_{Bn}/k_BT} \\ &= A^*T^2e^{-q\phi_{Bn}/k_BT} \end{aligned} \tag{4.25}$$

A^* 称为里查孙常数，

$$A^* = \frac{4\pi q m_n k_B^2}{h^3} \tag{4.26}$$

h 是普朗克常数。对于自由电子，里查孙常数 A_0^* 为 120A/(cm^2·K^2)。所以，对于导带能谷球对称(电子有效质量各向同性)的半导体，例如 GaAs，$A^*/A_0^* = m_n/m_0$，即

$$A^* = 120(m_n/m_0) \quad [\mathrm{A/cm^2 \cdot K^2}] \tag{4.27}$$

m_n 和 m_0 分别为导带能谷电子有效质量和自由电子质量。

方程(4.24)表明，金属-半导体接触和 p-n 结有完全类似的电流电压关系，但金属-半导体接触中电子热发射代替了 p-n 结中非平衡少数载流子的扩散，即 v_c 代替了 D/L(D 和 L 分别为载流子的扩散系数和扩散长度)，前者比后者要大几个数量级。即对同样的 V_{bi}，金属-半导体接触中的电流比 p-n 结大几个数量级；或者说，对同样的使用电流，金属-半导体接触有较低的导通电压。

除了多数载流子(电子)电流以外，在金属和 n 型半导体的接触中，由于金属向半导体注入空穴(实际上是半导体价带顶附近的电子流向金属中费米能级以下的空状态而在半导体中留下空穴)，电流中也有少数载流子(空穴)电流，密度为

$$J_p = \frac{qD_p p_{n0}}{L_p}(e^{qV/k_BT} - 1) \tag{4.28}$$

这个电流一般比电子(多子)电流小好几个数量级，故相应的器件(肖特基势垒二极管)常称为多数载流子器件。

在金属-半导体中存储效应可以忽略，原因是电流取决于多数载流子。以金属和 n 型半导体接触来说，正向电流主要是 n 型半导体中的电子热发射进入金属，它们进入金属以后成为漂移电流而流走，不存在电荷存储。存储电荷的积累或消失需要一个弛豫过程，将严重限制器件的速度和频率响应。换句话说，由于肖特基势垒二极管中没有电荷存储，它被广泛用于高速、高频电路。

在小信号应用中，未封装的肖特基势垒二极管可以用图 4.3 所示的等效电路表示。图中，C_j 为二极管的结电容，r_d 为二极管的结电阻，r_s 为串联电阻。

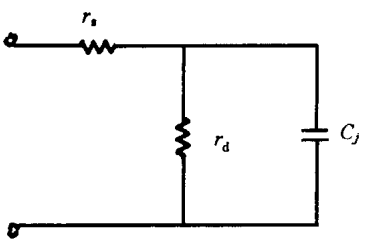


图 4.3 肖特基二极管的等效电路

$$C_j = A\sqrt{\frac{q\varepsilon_s N_D}{2(V_{bi} - V)}}$$

A 表示结的面积。

$$r_d = \frac{dV}{dI} = \frac{k_BT}{qI}$$

当二极管作为检波器或整流器使用时，要求交流功率主要被结电阻吸收，而在串联电阻上耗散很少。通常 $r_s \ll r_d$，所以低频时 r_s 的作用可以忽略。但是，随着使用频率 ω 的上升，结阻抗 $Z_d = (1/r_d + j\omega C_j)^{-1}$ 相对 r_s 降低，吸收的功率相对 r_s 上耗散的功率减少。定义截止频率为吸收功率和耗散功率相等时的工作频率，则应有

$$r_s = \frac{r_d}{1 + (\omega_T C_j r_d)^2} \tag{4.29}$$

ω_T 为截止圆频率。对于 $r_d \gg r_s$，有

$$\omega_T^2 \approx \frac{1}{C_j^2 r_d r_s} \tag{4.30}$$

由此可得截止频率 f_T，有

$$\frac{1}{2\pi C_j r_d} < f_T < \frac{1}{2\pi C_j r_s} \tag{4.31}$$

对于高频应用，要求 C_j、r_d 和 r_s 都很小。如果半导体掺杂浓度较高和载流子迁移率很大，则 r_s 可以做到很小。

4.1.3 欧姆接触

金属-半导体接触的另一个重要应用是作为器件引线的电极接触，这种接触不应当影响器件的电学特性，称为欧姆接触。所以，作为欧姆接触，金属和半导体接触的电阻应小到与半导体的体电阻相比可以忽略。表示欧姆接触性质的基本参数是接触电阻率（又称比接触电阻），我们以 ρ_c 表之，其定义为

$$\rho_c \equiv \left(\frac{\partial J}{\partial V}\right)^{-1}\Big|_{V\to 0} \quad [\Omega \cdot \mathrm{cm}^2] \tag{4.32}$$

对于低掺杂半导体，应用(4.24)式，可得

$$\rho_c = \frac{k_B}{qA^* T} e^{q\phi_{Bn}/k_B T} \tag{4.33}$$

对于高掺杂半导体的接触，势垒宽度 W 很小，电流输运主要是隧道穿透，电流与穿透几率成正比，

$$J \sim e^{-\frac{2}{\hbar}W\sqrt{2m_n(\phi_{Bn}-V)}} \tag{4.34}$$

用 $W \approx \sqrt{\frac{2\epsilon_s}{qN_D}(\phi_{Bn}-V)}$ 代入上式，并根据 ρ_c 的定义，可得

$$\rho_c \sim e^{C_1\phi_{Bn}/\sqrt{N_D}} \tag{4.35}$$

式中 $C_1 = 2\sqrt{m_n\epsilon_s}/\hbar$。式(4.35)说明，在隧道效应范围内，接触电阻率强烈依赖于掺杂浓度，且随因子 $\phi_{Bn}/\sqrt{N_D}$ 指数下降。

综上所述，为了获得低的 ρ_c，必须采用高掺杂浓度或低势垒高度，或二者都用。事实上，这两种方法是制作欧姆接触的基本方法。

4.2 GaAs MESFET

这类器件的工作原理和结型场效应晶体管(JFET)相同。不过，它用肖特基势垒代替了p-n结，通过其耗尽层厚度改变来调制电流通道（导电沟道）的横截面。与 JFET 相比，MESFET 在工艺制作和特性方面有某些优点。例如，金属-半导体接触势垒在低温下形成（而 p-n 结却要在高温下采用扩散或生长工艺制造），因而可以采用像 GaAs 这样电子迁移率高和饱和速度大的化合物半导体材料，而得到开关速度快和使用频率高的器件。

4.2.1 基本结构

n 沟 GaAs MESFET 是目前常见的器件，在微波和高速数字电路方面有许多重要应用。实际的 GaAs MESFET 是在半绝缘 GaAs 衬底上由离子注入或薄膜生长形成的 n 型 GaAs 薄

层上制作的。采用半绝缘衬底是为了减小寄生电容(沟道-衬底电容)和便于器件之间的隔离。此外,栅极金属一般用难熔金属(例如 TiW 合金)同时充当形成源、漏接触区的掩膜,使源、漏接触尽可能靠近栅极,以减小沟道两端的串联电阻。

图 4.4 表示这类 MESFET 的基本结构。使用时,源极一般接地,栅电压和漏电压都是相对源量度的。在正常工作条件下,漏极正向偏置,即 $V_{DS}>0$,因而肖特基结的耗尽区向漏端偏移;栅压 V_{GS}应小于肖特基结的正向导通电压以避免明显的栅极漏电流,对 n-GaAs MESFET,V_{GS}的最高极限约为 +0.7V。n-GaAs 层的掺杂浓度 N_D 约为 $10^{17}cm^{-3}$。器件的基本结构尺寸是沟道长度 L、沟道宽度 Z 和沟道深度 a。

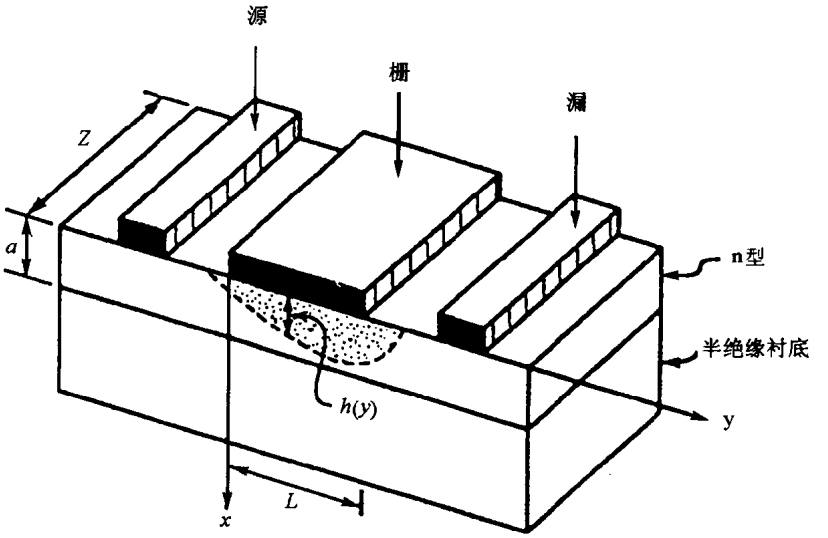


图 4.4 n-GaAs MESFET 的基本结构

4.2.2 夹断电压和阈值电压

设 n-GaAs 层均匀掺杂,则根据单边突变 p^+-n 结的理论,肖特基结的耗尽层厚度为

$$h(y) = \sqrt{\frac{2\varepsilon_s}{qN_D}[V_{bi} - V_{GS} + V(y)]} \tag{4.36}$$

式中,$V(y)$是沟道内 y 处相对源端的电势,在漏端,$V(y=L)=V_{DS}$;V_{bi}是肖特基势垒的自建电压;V_{GS}是栅压。y 处肖特基势垒上的总电势为 $V_{bi}-V_{GS}+V(y)$。当 $h=a$ 时,势垒耗尽层边界和衬底接触,相应的总电势称为夹断电压,以 V_p 表示,显然有

$$V_p = \frac{qN_D}{2\varepsilon_s}a^2 \tag{4.37}$$

忽略串联电阻的作用,则在沟道源端($y=0$)有 $V(y=0)=0$,使耗尽层边界和衬底接触的临界栅压(器件临界截止)显然为

$$V_T = V_{bi} - V_p \tag{4.38}$$

V_T 称为阈值电压。只有 V_{GS}超过 V_T,器件才会有导电沟道,而进入导通状态。如果半导体层(n-GaAs 层)的厚度 a 很小或者掺杂浓度 N_D 比较低,则 $V_G=0$ 时器件因自建电压 V_{bi}的作

用在源端夹断，而不存在导电沟道，只有在超过 V_T 的正栅压下才能导通。这类器件称为增强型(常断型)。相反，如果 a 较厚或者 N_D 较高，$V_G=0$ 时已经有导电沟道，则这类器件称为耗尽型(常开型)；耗尽型器件只有在比 V_T 更负的栅压下才会截止。

4.2.3 电流-电压特性

1. 恒定迁移率模型

假定漏电压 V_{DS}不很高，以致沿沟道方向(y 方向)的电场强度处处低于速度饱和电场，电流和电压的关系服从欧姆定律，则 y 处的电流密度可表示为

$$J(y) = qN_D\mu_n\mathscr{E}_y(y) \tag{4.39}$$

电子迁移率 μ_n 为常数。$\mathscr{E}_y(y)=\mathrm{d}V(y)/\mathrm{d}y$ 为离源端 y 处的场强数值，将其代入上式，可得漏电流 I_{DS}的表示式：

$$I_{DS} = qN_D\mu_n\left(\frac{\mathrm{d}V}{\mathrm{d}y}\right)[a - h(y)]Z \tag{4.40}$$

其中$[a-h(y)]Z$ 为沟道的横截面，$h(y)$为 y 处的耗尽层厚度，已由式(4.36)给出。

漏电流 I_{DS}是和 y 无关的常数，式(4.40)可改写为

$$I_{DS}\mathrm{d}y = q\mu_nN_DZ[a - h(y)]\mathrm{d}V \tag{4.41}$$

利用方程(4.36)可得

$$\mathrm{d}V = \frac{qN_D}{\varepsilon_s}h(y)\mathrm{d}h \tag{4.42}$$

将 $\mathrm{d}V$ 代入式(4.41)，并从 $y=0$ 到 $y=L$ 进行积分，得到

$$\begin{aligned} I_{DS} &= \frac{1}{L}\int_{h_1}^{h_2} q\mu_nN_DZ(a-h)\frac{qN_D}{\varepsilon_s}h\,\mathrm{d}h \\ &= \frac{Z\mu_nq^2N_D^2}{2\varepsilon_sL}\left[a(h_2^2-h_1^2) - \frac{2}{3}(h_2^3-h_1^3)\right] \end{aligned} \tag{4.43}$$

h_1 和 h_2 分别为沟道源端和漏端的耗尽层厚度，由式(4.26)可得

$$h_1 = \sqrt{\frac{2\varepsilon_s}{qN_D}(V_{bi} - V_{GS})} \tag{4.44}$$

$$h_2 = \sqrt{\frac{2\varepsilon_s}{qN_D}(V_{bi} - V_{GS} + V_{DS})} \tag{4.45}$$

将式(4.44)，(4.45)代入式(4.43)，并加整理后得

$$I_{DS} = I_p\left[\frac{V_{DS}}{V_p} - \frac{2}{3}\left(\frac{V_{DS} - V_{GS} + V_{bi}}{V_p}\right)^{3/2} + \frac{2}{3}\left(\frac{V_{bi} - V_{GS}}{V_p}\right)^{3/2}\right] \tag{4.46}$$

$$I_p = \frac{Z\mu_nq^2N_D^2a^3}{2\varepsilon_sL} \tag{4.47}$$

$$V_p = \frac{qN_Da^2}{2\varepsilon_s} \tag{4.48}$$

耗尽型 n-MESFET 的基本 I-V 特性如图 4.5 所示。对于不比 V_T 更负的栅源电压 V_{GS}，将有漏极电流 I_{DS}流动。增强型 n-MESFET 的 I-V 特性在形状上完全类似，但阈值电压 V_T 比 $0V$ 稍高。

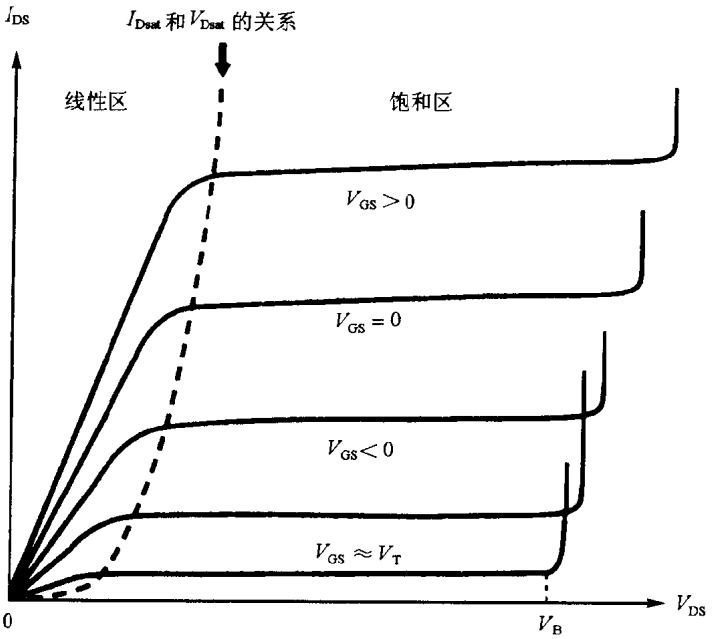


图 4.5　耗尽型 n-MESFET 的基本 *I-V* 特性

从图 4.5 可以看到，*I-V* 特性有两个不同的区域。当 V_{DS}很小时，沟道截面积(或者说沟道厚度 $a-h$)基本上与 V_{DS}无关，*I-V* 特性是欧姆性的或线性的，因而这个工作区通常称为线性区。当 $V_{DS}\geqslant V_{Dsat}$时，电流达到饱和值 I_{Dsat}，这一工作区称为饱和区。随着 V_{DS}进一步增加，最终导致栅-沟道二极管发生雪崩击穿，这时 I_{DS}突然增加。

线性区　在这一工作区，$V_{DS}\ll V_{bi}-V_{GS}$，将式(4.46)展开，可得

$$I_{DS}\approx \frac{I_p}{V_p}\left[1-\left(\frac{V_{bi}-V_{GS}}{V_p}\right)^{1/2}\right]V_{DS}$$

$$=\frac{q\mu_n N_D Za}{L}\left[1-\left(\frac{V_{bi}-V_{GS}}{V_p}\right)^{1/2}\right]V_{DS} \tag{4.49}$$

沟道电导(也叫漏电导)g_D 可表示为

$$g_D=\left.\frac{\partial I_{DS}}{\partial V_{DS}}\right|_{V_{GS}}=\frac{q\mu_n N_D Za}{L}\left[1-\left(\frac{V_{bi}-V_{GS}}{V_p}\right)^{1/2}\right] \tag{4.50}$$

我们注意到，方括号内的因子实际上表示沟道厚度 $a-h$，即

$$a-h=a\left[1-\left(\frac{V_{bi}-V_{GS}}{V_p}\right)^{1/2}\right] \tag{4.51}$$

它和漏电压 V_{DS}无关。这时从源到漏形成近似均匀的导电沟道，其厚度随栅压 V_{GS}的增加而增加。

MESFET 的另一个重要参数是跨导 g_m，其定义为

$$g_m\equiv\left.\frac{\partial I_{DS}}{\partial V_{GS}}\right|_{V_{DS}} \tag{4.52}$$

跨导表示在给定的漏电压下，栅电压改变引起的漏电流的变化。在线性区，g_m 可由式(4.49)

得到：

$$g_m = \frac{I_p}{\partial V_p^2}\left(\frac{V_p}{V_{bi} - V_{GS}}\right)^{1/2} V_{DS} = \frac{Z\mu_n}{L}\left[\frac{qN_D\varepsilon_s}{2(V_{bi} - V_{GS})}\right]^{1/2} V_{DS} \tag{4.53}$$

饱和区 在式(4.46)中令 $V_{bi} - V_{GS} + V_{DS} = V_p$，就得到夹断点电流，亦即饱和区漏电流，我们以 I_{Dsat}表示之，显然有

$$I_{Dsat} = I_p\left[\frac{1}{3} - \frac{V_{bi} - V_{GS}}{V_p} + \frac{2}{3}\left(\frac{V_{bi} - V_{GS}}{V_p}\right)^{3/2}\right] \tag{4.54}$$

相应的饱和电压是

$$V_{Dsat} = V_p - V_{bi} + V_{GS} = V_{GS} - V_T \tag{4.55}$$

在理想情形下，由于式(4.55)中 I_{Dsat}不是 V_{DS}的函数，饱和区的沟道电导是零；实际上，随 V_{DS}增加，夹断点从漏端向源端移动，有效沟道长度缩短，饱和区有非零的沟道电导。饱和区的跨导可根据其定义和式(4.54)得到：

$$g_m = \frac{Z\mu_n qN_D a}{L}\left[1 - \left(\frac{V_{bi} - V_{GS}}{V_p}\right)^{1/2}\right] \tag{4.56}$$

雪崩击穿 发生在沟道漏端，这里栅-沟道结的反向电压最高。这时漏电压 V_{DS}和反向击穿电压 V_B 之间有如下关系：

$$V_B = V_{DS} - V_{GS} \tag{4.57}$$

2. 饱和速度模型

目前，场效应晶体管(FET)的沟道长度已经减小到 1μm 以下，沿沟道的电场通常很强，以致漏极电流饱和并非由于沟道夹断，而是因为电子漂移速度达到饱和。实验表明，对于栅长在 0.5～2μm 的短沟道 GaAs MESFET，速度饱和模型能够精确地描述饱和区的电流特性。图 4.6可以用来说明和推导这一模型，该图表示工作在电流饱和区的 GaAs MESFET 沟道截面，最窄的沟道缝隙出现于沟道区的漏端。在饱和区，栅耗尽层下的沟道电场显著高于电子谷间转移的临界电场(对 GaAs 约等于 3.5kV/cm)，从源向漏行进的电子将从高迁移率的低能谷(Γ 能谷)散射到低迁移率的高能谷(L 能谷)，平均速度在接近漏端的区域下降并且有效饱和。为了维持电流的连续性，在该区发生了电子的强积累，这是由于随离开源的距离增加沟道缝隙逐渐变窄且电子漂移速度逐渐下降的缘故。离开漏端的情况恰恰相反，由于沟道展宽并且电子运动加快，产生了一个因为电子部分耗尽所致的强耗尽层。积累层和耗尽层的电荷接近相等，大部分漏电压降落在这一静止不动的偶极层上。在其他像 GaAs 一样具有低场高迁移率的材料(例如 InGaAS)制作的 FET 中，漏极电流饱和的机制相同。Si、SiC 和金刚石等材料制作的 FET 也有类似的漏极电流饱和现象，但主要来自沟道的贡献。

根据上面的分析，可以假定沟道电子以恒定的有效饱和速度 v_s 运动(v_s 的典型值约为 1.2×10^7cm/s)，从而饱和电流直接受耗尽层厚度 h 和沟道深度 a 之差调制。这就是速度饱和模型。对于均匀掺杂材料，由这一模型得到

$$I_{DS} = qv_s ZN_D(a - h) \tag{4.58}$$

用公式(4.51)替换 $a - h$，上式变为

$$I_{DS} = qv_s ZN_D a\left[1 - \left(\frac{V_{bi} - V_{GS}}{V_p}\right)^{1/2}\right] \tag{4.59}$$

上式对 V_{GS}微商，并利用 $V_p = qN_Da^2/2\varepsilon_s$，即得器件本征跨导的表示式

$$g_m = v_s Z\left[\frac{qN_D\varepsilon_s}{2(V_{bi} - V_{GS})}\right]^{1/2} \tag{4.60}$$

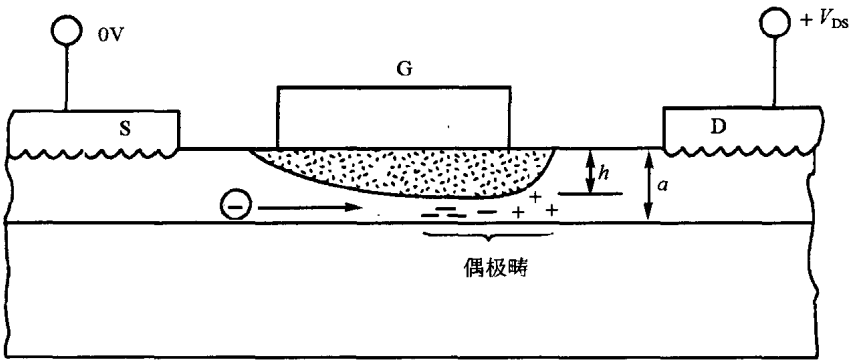


图 4.6　工作在电流饱和区的 GaAs MESFET 的沟道截面；在电子速度饱和的沟道段，栅耗尽层厚度 h 是常数

对于栅长明显小于 0.5μm 的 GaAs FET，电子通过栅下高场区的时间很短，以致来不及发生谷间转移，只受到轻微的声子散射、杂质或其他电子的散射，它们近似弹道运动。也就是说，电子在渡越沟道期间得到的平均动能几乎等于它所经历的电势降落，平均速度可以达到稳态饱和速度的若干倍。这就是所谓的速度过冲。在其他 Ⅲ-V 族的短沟道 FET 中，速度过冲也是普遍现象。但在短沟道 Si 器件中电子速度增加很少，原因是 Si 的能量弛豫时间总体上比 GaAs 的约小一个数量级，对于电子能够漂移一段宏观的距离来说还是太短，所以没有明显的电子速度过冲效应。

4.2.4　截止频率

当栅压变化时，一部分沟道电流要对栅电容充电。截止频率 f_T 定义为该频率下器件不再放大输入信号，这时输入栅电容的电流将等于漏端输出电流；即 f_T 是 MESFET 的短路电流增益下降为 1 时的频率。我们可以通过如图 4.7 所示的简化等效电路来得到 f_T 的表达式。

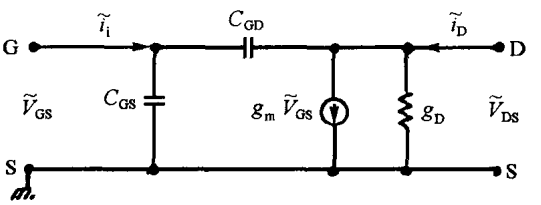


图 4.7　MESFET 的简单等效电路

根据图 4.7，输入电流 $\tilde{i}_i$ 和输出电流 $\tilde{i}_D$ 分别为

$$\tilde{i}_i = 2\pi f(C_{GS} + C_{GD})\tilde{V}_{GS} = 2\pi f C_G\tilde{V}_{GS} \tag{4.61}$$

$$\tilde{i}_D = g_m\tilde{V}_{GS} \tag{4.62}$$

令 $\tilde{i}_D/\tilde{i}_i=1$，得到

$$f_T = \frac{g_m}{2\pi C_G} \tag{4.63}$$

其中 $C_G = C_{GS} + C_{GD}$，C_{GS}是栅和源之间的结电容，C_{GD}是栅和漏之间的结电容。设栅-沟道二极管的平均耗尽层厚度为$a/2$，则栅电容 C_G 可表示为

$$C_G = LZ\frac{\varepsilon_s}{a/2} = \frac{2LZ\varepsilon_s}{a} \tag{4.64}$$

由式(4.63)可知，要提高截止频率，应该有大的跨导 g_m 和小的栅电容 C_G，也就是应采用高载流子迁移率和短沟道长度的 MESFET。

截止频率 f_T 同栅极下面电子的渡越时间 τ 有直接的关系。我们从器件的跨导 g_m 和栅电容 C_G 出发来讨论这种关系，因为 f_T 由 g_m 和 C_G 决定。根据定义，有

$$g_m = \frac{\Delta I_{DS}}{\Delta V_G} \tag{4.65}$$

$$C_G = \frac{\Delta Q_G}{\Delta V_G} \tag{4.66}$$

假定栅压有一小的增量 ΔV_G，则它将引起下述变化：

(1) 栅极作为栅-沟道电容器的一个“极板”，其上的栅电荷增加 $\Delta Q_G = C_G \times \Delta V_G$；

(2) 沟道作为电容器的另一个“极板”，沟道电荷将有一个负的增量 $-\Delta Q_G$(即电子浓度有一正的增量)，从而导致通过沟道的电流增加 ΔI_{DS}；而 ΔI_{DS}等于电荷增量除以它通过距离 L_g(栅长)的渡越时间 τ，即

$$\Delta I_{DS} = \frac{\Delta Q_G}{\tau} \tag{4.67}$$

利用式(4.66)表示 ΔQ_G，有

$$\Delta I_{DS} = V_G \times \frac{\Delta V_G}{\tau} \tag{4.68}$$

进一步，从式(4.65)得到

$$\frac{g_m}{C_G} = \frac{1}{\tau} \tag{4.69}$$

将式(4.69)代入式(4.63)，我们得到截止频率 f_T 和渡越时间 τ(或 L_g/v_s，设载流子以饱和速度 v_s 运动)之间关系的表达式：

$$f_T = \frac{1}{2\pi\tau} \approx \frac{v_s}{2\pi L_g} \tag{4.70}$$

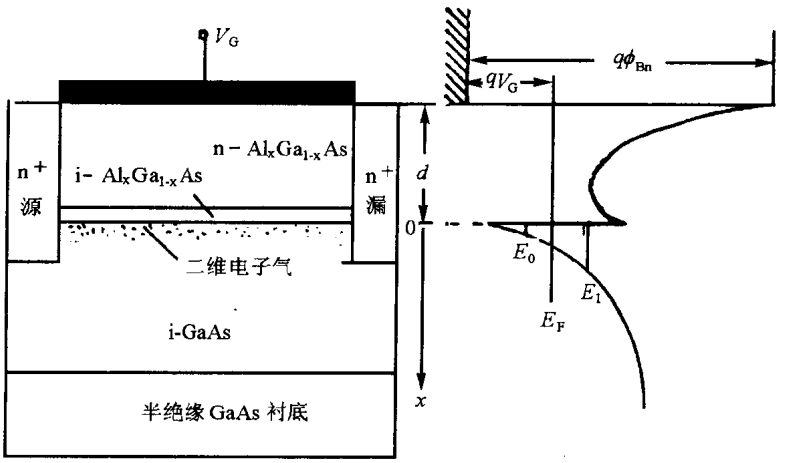
公式(4.70)表明，简单地通过计算电子在栅极下面的渡越时间，就可以判断不同场效应晶体管的高频性能。它也表明，重要的是实际栅区的长度，而不是从源至漏的整个距离。

4.3 高电子迁移率晶体管(HEMT)

4.3.1 器件结构及特点

这是一种异质结 MESFET。图 4.8 表示 n 沟道 AlGaAs/GaAs HEMT 的基本结构和能

带。对于这类器件，通过对栅极施加正向偏压，可以将电子引入异质结界面处的 GaAs 中。GaAs 不掺杂。AlGaAs 层通常称为控制层，它和金属栅极形成肖特基势垒结，和 GaAs 层形成异质结。AlGaAs 层的厚度和掺杂浓度(典型值分别为数百 Å 和 $10^{17}\sim10^{18}cm^{-3}$)决定器件的阈值电压，正常情形下应使之完全耗尽。如果 AlGaAs 层较厚或者掺杂浓度较高，则栅压 $V_G=0$ 时，异质结界面处 GaAs 表面的电子势阱内已经有电子存在，HEMT 是耗尽型的；相反，AlGaAs 层较薄或者掺杂浓度较低，$V_G=0$ 时耗尽层伸展到 GaAs 层内部，势阱内没有电子，器件是增强型的。AlGaAs 的禁带宽度比 GaAs 的大，它们形成异质结时，导带边不连续，AlGaAs 的导带边比 GaAs 的要高 ΔE_c，实际上就是前者的电子亲和能比后者的小，结果电子从 AlGaAs 向 GaAs 中转移，引起界面处能带弯曲，在 GaAs 表面形成近似三角形的电子势阱。当势阱较深时，电子基本上被限制在势阱宽度所决定的薄层(~100Å)内，称之为二维电子气(2DEG)。2DEG 是指电子(或空穴)在平行于界面的平面内自由运动，而在垂直于界面的方向受到限制。势阱深度受到加在栅极上的电压控制，故 2DEG 的浓度(面密度)将受 V_G 控制，从而器件的电流受 V_G 控制。



(a) 基本结构　　(b) 能带图

图 4.8　AlGaAs/GaAs HEMT

2DEG 和提供自由电子的 n^+-AlGaAs 层是空间分离的，中间还隔着一个不掺杂的 i-AlGaAs 薄层(20－70Å)，从而基本上不受电离杂质散射，迁移率显著增加，比体材料的电子迁移率高得多(为了避免势阱内散射，GaAs 层也是不掺杂的)。HEMT 依靠迁移率很高的 2DEG 导电，因此比起 GaAs MESFET 来有更大的工作速度和更高的截止频率。

HEMT 又称 2DEG 场效应晶体管(TEGFET)，或调制掺杂场效应晶体管(MODFET)，选择掺杂异质结晶体管(SDHT)等。前者说明这类器件高速高频的原因是电子迁移率高，后者强调得到高电子迁移率的方法是调制掺杂或者说选择掺杂。

4.3.2　2DEG 浓度(面密度)

选取如图 4.8 所示的坐标，则 2DEG 的电子在 x 方向的运动受到限制。对于无限深三角形势阱近似，能量本征值为

$$E_i = \left(\frac{\hbar^2}{2m_x}\right)^{1/3}\left[\frac{3}{2}\pi q\mathscr{E}\right]^{2/3}\left(i+\frac{3}{4}\right)^{2/3} \tag{4.71}$$

$$i = 0,1,2,\cdots$$

$\hbar = h/2\pi$ 是约化普朗克常数，$\mathscr{E}$ 是势阱中的平均电场。在与 x 方向垂直的 $y-z$ 平面内电子自由运动，能量连续变化，可表示为

$$E_\perp = \frac{\hbar^2}{2}\left(\frac{k_y^2}{m_y}+\frac{k_z^2}{m_z}\right) \tag{4.72}$$

m_x, m_y, m_z 分别是电子沿 x, y, z 方向的有效质量；k_y 和 k_z 分别是 y 方向和 z 方向的波矢。电子在势阱中运动的总能量为

$$E = E_i + E_\perp \tag{4.73}$$

从式(4.73)可以看出，2DEG 的能级形成许多个子能带，每个子能带具有最小能量 E_i。

对于单纯的二维运动，k 空间的电子等能面退化为一个椭圆，仿照 1.1 节关于态密度的讨论，可以证明相应的能态密度为

$$D(E) = \frac{\sqrt{m_y m_z}}{\pi\hbar^2} \tag{4.74}$$

这是与能量无关的常数。若有效质量各向同性(球形等能面情形)，$m_x = m_y = m_z = m$，则有

$$D(E) = \frac{m}{\pi\hbar^2} \tag{4.75}$$

对于准二维运动，能量低于 E_i 的状态不存在，能量高于 E_i 是二维运动。这样，温度为 T 时，第 i 个子能带中的电子数(面密度)为

$$N_i = \int_{E_i}^{\infty} D(E)f(E)\mathrm{d}E = \frac{m}{\pi\hbar^2}\int_{E_i}^{\infty}\left[1+\mathrm{e}^{(E-E_\mathrm{F})/k_\mathrm{B}T}\right]^{-1}\mathrm{d}E$$

$$= \frac{mk_\mathrm{B}T}{\pi\hbar^2}\ln\left[1+\mathrm{e}^{(E_\mathrm{F}-E_i)/k_\mathrm{B}T}\right] \tag{4.76}$$

对于 E_i 有 l 个允许值的 2DEG 系统，总的电子数(面密度)为

$$N_\mathrm{s} = \sum_{i=0}^{l} N_i \tag{4.77}$$

势阱中的电场强度 $\mathscr{E}$ 可以用势阱两个边界处场强的平均值来表示，即

$$\mathscr{E} = \frac{1}{2}\left[q(N_\mathrm{s}+N_\mathrm{B})/\varepsilon_1 + qN_\mathrm{B}/\varepsilon_1\right] \tag{4.78}$$

为具体起见，考虑 AlGaAs/GaAs 结构。这时，势阱在 GaAs 内，$q(N_\mathrm{s}+N_\mathrm{B})/\varepsilon_1$ 是异质结边界处的电场强度，$qN_\mathrm{B}/\varepsilon_1$ 是势阱与 GaAs 衬底交界面处的电场强度，ε_1 和 N_B 分别为 GaAs 的介电常数和掺杂浓度。由于 GaAs 不掺杂，$N_\mathrm{B}\simeq 0$，所以势阱内的平均电场为

$$\mathscr{E} \simeq qN_\mathrm{s}/2\varepsilon_1 \tag{4.79}$$

利用式(4.79)可将表示 E_i 的公式(4.71)改写为

$$E_i = \left(\frac{9\hbar^2q^4}{128m\varepsilon_1^2}\right)^{1/3}\left[i+\frac{3}{4}\right]^{2/3}N_\mathrm{s}^{2/3}$$

$$= \gamma_i N_\mathrm{s}^{2/3} \tag{4.80}$$

式中，

$$\gamma_i = \left(\frac{9h^2q^4}{128m\varepsilon_1^2}\right)^{1/3}\left[i+\frac{3}{4}\right]^{2/3} \tag{4.81}$$

h 是普朗克常数，q 是电子电荷。对于 GaAs 中的电子，$m=0.067m_0$（m_0 是自由电子质量），$\varepsilon_1 \simeq 1.15\times10^{-14}$F/cm，故有

$$\gamma_i \simeq 2.1\times10^{-12}\left(i+\frac{3}{4}\right)^{2/3}[\text{eV}\cdot\text{m}^{4/3}] \tag{4.82}$$

在实际器件中，通常只需要考虑二个最低的子能带，即 $i=0,1$ 的情形，这时 2DEG 的浓度为

$$N_s = \frac{mk_BT}{\pi\hbar^2}\ln[(1+e^{(E_F-E_0)/k_BT})(1+e^{(E_F-E_1)/k_BT})] \tag{4.83}$$

其中

$$E_0 = \gamma_0 N_s^{2/3},\quad E_1 = \gamma_1 N_s^{2/3} \tag{4.84}$$

对于 GaAs，根据实测结果，$\gamma_0 \simeq 2.5\times10^{-12}\text{eV}\cdot\text{m}^{4/3}$，$\gamma_1 \simeq 3.2\times10^{-12}\text{eV}\cdot\text{m}^{4/3}$（理论计算分别为 1.7×10^{-12}和 3.0×10^{-12}）。

把式(4.84)代入式(4.83)中，就得到 2DEG 的费米势 E_F/q 和它的浓度 N_s 之间的函数关系。计算结果表明，N_s 之值在 $5\times10^{11}\text{cm}^{-2}$和 $1.5\times10^{12}\text{cm}^{-2}$之间时，$E_F/q$ 和 N_s 有线性关系：

$$\frac{E_F}{q} = \frac{E_{F0}}{q} + aN_s \tag{4.85}$$

$T=300$K 时，$E_{F0}\simeq 0$，$a\simeq 0.125\times10^{-6}\text{V}\cdot\text{m}^2$。

4.3.3 HEMT 的基本性质

1. 电荷和电压的关系

由图 4.8(b)给出的能量关系不难得到下述关于电压的方程：

$$\phi_{Bn} - V_G = -\left(\frac{qN_s}{\varepsilon_2}d - \frac{q}{\varepsilon_2}\int_0^{-d}N_D(x)x\,dx\right) + \frac{\Delta E_c}{q} - \frac{E_F}{q} \tag{4.86}$$

ϕ_{Bn}是金属接触的势垒高度，V_G 是栅压，ΔE_c 是异质结材料（这里为 AlGaAs 和 GaAs）导带边的能量差，E_F/q 是 2DEG 的费米势，ε_2 是控制层材料(AlGaAs)的介电常数。上式右边括号内表示的项是控制层上的电压降，其中$(qN_s/\varepsilon_2)d$ 由 2DEG 引起，余者则由电离杂质产生。当控制层均匀掺杂时，$N_D(x)=$常数，电离杂质产生的电压降为$(qN_D/2\varepsilon_2)d^2$。将 $E_F/q = E_{F0}/q + aN_s$ 代入上式，并进行整理，可得

$$qN_s = \frac{\varepsilon_2}{d+\Delta d}(V_G - V_T) \tag{4.87}$$

其中

$$V_T = \phi_{Bn} - \frac{qN_D}{2\varepsilon_2}d^2 - \frac{\Delta E_c}{q} + \frac{\Delta E_{F0}}{q} \tag{4.88}$$

$$\Delta d = \frac{\varepsilon_2 a}{q} \simeq 80\text{Å} \tag{4.89}$$

V_T 是 HEMT 的阈值电压，Δd 是 2DEG 的有效厚度。

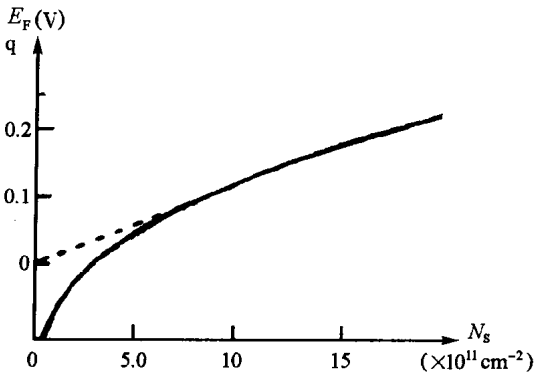


图 4.9　300K 时 2DEG 的费米势 E_F/q 及其浓度 N_s 之间的关系

当漏、源间加电压 V_{DS}时，2DEG 薄层构成导电沟道。选择沿沟道(从源至漏)的方向为 y 方向，源端为 $y=0$。以 $V(y)$表示沟道内 y 处的电位，则控制电荷薄层的有效栅压为 $V_{GS}-V(y)$，表示电荷与电压关系的方程(4.87)这时应变成

$$qN_s(y) = \frac{\varepsilon_2}{d+\Delta d}[V_{GS} - V_T - V(y)] \tag{4.90}$$

或

$$qN_s(y) = qN_{s0} - \frac{\varepsilon_2}{d+\Delta d}V(y) \tag{4.91}$$

N_{s0}是源端的电荷浓度。

2. 电流和电压的关系

对于长沟道器件，沿沟道方向的电场处处低于速度饱和电场，沟道电流可表示为

$$\begin{aligned} I_{DS} &= ZqN_s(y)\mu\mathscr{E}(y) \\ &= \mu Z\frac{\varepsilon_2}{d+\Delta d}[V_{GS} - V_T - V(y)]\frac{dV(y)}{dy} \end{aligned} \tag{4.92}$$

Z 为沟道宽度，μ 为低场电子迁移率，$\mathscr{E}(y)=\dfrac{dV(y)}{dy}$是 y 处沟道电场强度。将上式从沟道的源端($y=0$)至漏端($y=L$，L 是沟道长度)积分，得到

$$I_{DS} = \mu\left(\frac{Z}{L}\right)\frac{\varepsilon_2}{d+\Delta d}[(V_{GS} - V_T)V_{DS} - V_{DS}^2/2] \tag{4.93}$$

当 V_{DS}增加到 $V_{Dsat}=V_{GS}-V_T$，由式(4.90)可知，$qN_s(L)=0$，即沟道在漏端夹断，电流达到饱和。在饱和区，漏极电流为

$$I_{Dsat} = \frac{\mu\varepsilon_2 Z}{2(d+\Delta d)L}(V_{GS} - V_T)^2 \tag{4.94}$$

对于短沟道器件($L\sim1\mu$m)，强电场下电子将以饱和速度 v_s 运动，漏极电流可表示为

$$\begin{aligned} I_{Dsat} &= qN_{s0}v_sZ \\ &= \frac{Z\varepsilon_2 v_s}{(d+\Delta d)}(V_{GS} - V_T) \end{aligned} \tag{4.95}$$

如前节所述，GaAS 中电子的稳态饱和速度 $v_s \widetilde{>} 1\times10^7\text{cm/s}$。对沟道十分短的器件，输运往往是“非稳态过程”，电子速度可以是稳态饱和速度的若干倍。

根据定义，通过 I_{DS}对 V_{GS}的微商（V_{DS}维持不变），可以求得器件的跨导 g_m。对于 I_{DS}非饱和的情形，由方程(4.93)可得

$$g_m = \mu\left(\frac{Z}{L}\right)\frac{\varepsilon_2}{d+\Delta d}V_{DS} \tag{4.96}$$

而在电流饱和区，由方程(4.94)和(4.95)，有

$$g_m = \begin{cases} \dfrac{Z\varepsilon_2\mu}{(d+\Delta d)L}(V_{GS}-V_T) & \text{（长沟道器件）} \\ \dfrac{Z\varepsilon_2 v_s}{(d+\Delta d)} & \text{（短沟道器件）} \end{cases} \tag{4.97}$$

3. 小信号栅电容

为了确定器件的小信号栅电容，必须首先计算栅极下面总的电荷量 Q_n，这可由下式得出

$$Q_n = Z\int_0^L(-qN_s)\mathrm{d}y = Z\int_0^{V_{DS}}(-qN_s)\frac{\mathrm{d}y}{\mathrm{d}V}\mathrm{d}V \tag{4.98}$$

$\frac{\mathrm{d}y}{\mathrm{d}V}$从式(4.92)得到为

$$\frac{\mathrm{d}y}{\mathrm{d}V} = \frac{ZqN_s\mu}{I_{DS}} \tag{4.99}$$

所以，

$$Q_n = -\frac{\mu Z^2}{I_{DS}}\int_0^{V_{DS}}(qN_s)^2\mathrm{d}V \tag{4.100}$$

将式(4.90)和(4.93)代入上式，可得

$$Q_n = \frac{2}{3}C_G\frac{V_{GD}'^3 - V_{GS}'^3}{V_{GS}'^2 - V_{GD}'^2} \tag{4.101}$$

式中

$$V_{GS}' = V_{GS}-V_T,\ V_{GD}' = V_{GS}'-V_{DS},\ C_G = \frac{\varepsilon_2 ZL}{d+\Delta d} \tag{4.102}$$

栅-源电容为

$$C_{GS} = -\frac{\partial Q_n}{\partial V_{GS}'} = \frac{2}{3}C_G\left[1-\frac{(V_{GS}'-V_{DS})^2}{(2V_{GS}'-V_{DS})^2}\right] \tag{4.103}$$

栅-漏电容为

$$C_{GD} = -\frac{\partial Q_n}{\partial V_{GD}'} = \frac{2}{3}C_G\left[1-\frac{V_{GS}'^2}{(2V_{GS}'-V_{DS})^2}\right] \tag{4.104}$$

这些电容都是在电流低于饱和值时得到的。当 $V_{DS}\to0$ 时，有

$$C_{GS} = C_{GD} = C_G/2 \tag{4.105}$$

当 $V_{DS} = V_{GS}' = V_{GS}-V_T$ 时，器件的电流特性开始饱和，这时有

$$C_{GS} = \frac{2}{3}C_G,\quad C_{GD} = 0 \tag{4.106}$$

4. 截止频率 f_T

与 MESFET 及其他 FET 一样，HEMT 的截止频率可以表示为

$$f_T = \frac{g_m}{2\pi C_G} \tag{4.107}$$

C_G 是器件本征部分的栅电容。如果把有效负载电容(边缘效应等引起的寄生电容也可以算在其内)C_L 也包括进来，则截止频率应表示成

$$f_T = \frac{g_m}{2\pi(C_G + C_L)} \tag{4.108}$$

或利用比值 g_m/C_G 和栅下电子渡越时间 τ 的关系(见式 4.69)，则这一表示实际器件截止频率的公式又可表为

$$f_T = \frac{1}{2\pi\tau} \cdot \frac{1}{1 + C_L/C_G} \tag{4.109}$$

对于 HEMT，本征栅电容为

$$C_G = \frac{\varepsilon_2 ZL}{d + \Delta d}$$

由于 d 可以做得比较小，C_G(即 g_m)较大，器件有较高的截止频率和较快的工作速度。

第五章　MOS 器件

与 JFET 或 MESFET 不同，栅压控制 MOS 器件导电沟道的载流子浓度。本章首先讨论 MOS 结构(MOS 二极管)的半导体表面状态，然后在此基础上讨论 MOS 场效应晶体管(MOSFET)。以这种器件为核心的 MOS 集成电路是当前超大规模集成电路(VLSI)的主流。电荷耦合器件(CCD)的基本结构是一种密排的 MOS 二极管阵列，被用于摄像、信息处理和数字存储等方面，在本章的最后介绍。

5.1　MOS 结构的基本性质

5.1.1　基本结构和能带图

MOS 结构指金属-氧化物-半导体结构。半导体作为衬底，我们主要考虑 p 型，并且假定掺杂均匀；氧化物常用 SiO_2，在 Si 上生长 SiO_2 的工艺简单可靠，SiO_2/Si 的界面态密度已能做到低于 $10^{10}cm^{-2}$；金属泛指栅极材料，不只限于金属，目前主要采用多晶硅或难熔金属硅化物。

图 5.1(a)是互相绝缘的金属、半导体及夹在中间的氧化层所组成系统的能带图；图 5.1(b)是热平衡时 MOS 二极管的能带图，它是对于金属和半导体之间有功函数差但 SiO_2 中不存在电荷的情形画的。对于图 5.1 所示的情形，金属的功函数 $q\phi_m$ 比半导体的功函数 $q\phi_s$ 小，故有电子离开金属，通过外电路或比 SiO_2 要好的通道向半导体转移。这样，栅极带正电，并排斥半导体表面的空穴使之形成由电离受主组成的空间电荷区。因为半导体表面处的空穴比远离

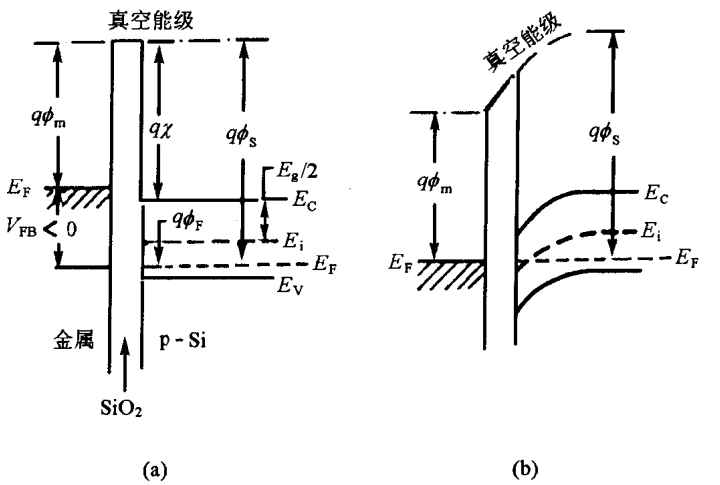


图 5.1　(a) 由互相绝缘的金属、半导体及夹在中间的氧化层所组成系统的能带图；(b) 热平衡时 MOS 二极管的能带图

表面的体内少，在表面处半导体的能带向下弯曲。由金属带正电、半导体表面带负电而产生的电位差分开降落在 SiO_2 层上和半导体表面空间电荷区上，这一电位差就是金属和半导体接触的电势差，它与金属和半导体之间的等效功函数（功函数除以基本电荷量 q）差数值相等但符号相反。对于我们现在讨论的情形，这一电势差大于零。

5.1.2 平带电压

前面的叙述说明，在 MOS 结构中，金属和半导体之间因功函数差而产生一定的固有电压，并造成半导体表面能带弯曲。可以想象，如果金属对半导体加相反电压使之平衡其固有电压，则半导体表面和体内一样，能带处处平坦。外加的能拉平半导体能带的电压就叫做平带电压，常用 V_{FB} 表示。对于实际的 MOS 结构，在 SiO_2-Si 系统中有所谓有效界面电荷，它将在金属和半导体内感应极性相反的电荷，是造成半导体能带不平的另一个原因。所以金属对半导体还必须再加一个电压，才能使半导体中的电荷完全消失，能带处处拉平。这就是说，对实际的 MOS 结构，平带电压可分成二部分电压相加，即 $V_{FB} = V_{FB1} + V_{FB2}$，$V_{FB1}$ 用来抵消功函数差的影响，V_{FB2} 用来消除有效界面电荷的影响。

（1）V_{FB1}　根据前面的讨论，这一电压等于金属和半导体的等效功函数差，

$$V_{FB1} = \frac{q\phi_m - q\phi_s}{q} = \phi_{ms} \tag{5.1}$$

从图 5.1 得出 $q\phi_s$，代入上式即有

$$V_{FB1} = \phi_G - \phi_F \tag{5.2}$$

其中

$$\phi_G = \frac{(q\chi + E_g/2) - q\phi_m}{(-q)} \tag{5.3}$$

$$\phi_F = \frac{E_F - E_i}{(-q)} \tag{5.4}$$

ϕ_F 是相对于本征费米能级 E_i 定义的半导体材料的费米势（ϕ_G 也有类似的意义），对于非简并的 p 型半导体有

$$\phi_F = \frac{k_B T}{q}\ln\left(\frac{N_A}{n_i}\right) \tag{5.5}$$

N_A 是受主掺杂浓度，n_i 是本征载流子浓度。对于给定的 MOS 结构，可由式(5.3)和(5.5)求出 V_{FB1}（即 ϕ_{ms}），ϕ_{ms} 决定于 MOS 结构所用的栅材料和半导体的掺杂浓度。

多晶硅是一种十分重要的栅材料。与金属栅相比，它的主要优点是能够承受器件制造中的高温过程。因此，多晶硅栅又可充当形成源、漏区的掩膜，得到没有栅源交叠或栅漏交叠的自对准栅。对于多晶硅栅（简称硅栅），式(5.2)中的 ϕ_G 应以多晶硅的费米势表示。由于作为栅材料，多晶硅通常是高掺杂的，费米能级靠近价带顶或导带底，即 $(E_F - E_i) \simeq \mp 0.56\text{eV}$，故有

$$\phi_G \simeq \pm 0.56V \tag{5.6}$$

“+”号用于 p 型栅，“−”号用于 n 型栅。

（2）V_{FB2}　金属和半导体表面出现电荷的另一个原因是 SiO_2 层内部以及 SiO_2/Si 界面存在电荷，它们是：界面陷阱电荷，氧化物固定电荷，氧化物陷阱电荷和可动离子电荷。界面陷

阱电荷 Q_{it}归因于 SiO_2/Si 界面性质，并取决于该界面的化学组分。在 SiO_2/Si 界面上的陷阱，其能级位于硅禁带之内。界面态密度(即单位面积界面陷阱数)和晶面取向有关，在〈100〉面(这里指法线方向为〈100〉的晶面，即|100|面。下同。在立方晶格中，晶面的密勒指数和晶面法线的晶向指数相同。)比〈111〉面的约小一个数量级。目前采用在 Si 上热生长 SiO_2 的 MOS 结构，其大部分界面陷阱电荷都用低温(450℃)氢退火方法中和掉了。对于硅〈100〉面，Q_{it}可以低到 $10^{10}cm^{-2}$，即大约每 10^5 个表面原子才有一个界面陷阱电荷。氧化物固定电荷 Q_f 位于SiO_2/Si界面层约 30Å 的范围内，这些电荷是固定的，基本上不随外加电压改变。Q_f通常是正的，并和氧化、退火条件及 Si 的界面取向有关，被认为是来自表面附近的硅离子并可看成是位于 SiO_2/Si 界面的电荷层。经过仔细处理的 SiO_2-Si 系统，Q_f 的典型值为 $10^{10}\ cm^{-2}$。氧化物陷阱电荷 Q_{ot}和 SiO_2 中的缺陷有关，例如在受到 X 射线辐射或高能电子轰击时就可能产生这类电荷。Q_{ot}大都可以通过低温退火消除。可动离子电荷 Q_m 来自工艺过程中诸如钠离子等碱金属离子玷污，它们在大偏压和高温条件下能在氧化层内移动，影响器件的可靠性，要特别注意防止。上述各类电荷是指单位面积有效的净电荷(以 C/cm^2 为单位)，现在来计算它们对平带电压的影响。为了简化分析，常常假定它们都固定在 SiO_2/Si 界面上，其面密度为 Q_0。对于 SiO_2-Si 系统，无论 p 型衬底或 n 型衬底，Q_0 总是正的，如上所述，其数值在现代工艺水平下可低至 $10^{-10}C/cm^2$。Q_0 将在金属和半导体中感应极性相反的电荷，感生电荷总量为($-Q_0$)。为了不使它对半导体产生影响，必须在金属上提供全部所需的 $-Q_0$，即除了 ϕ_{ms}以外，还应在外电路中接电源电压：

$$V_{FB2} = -Q_0/C_{ox} \tag{5.7}$$

电源的负极与金属相连，大小为 Q_0/C_{ox}。C_{ox}是栅 SiO_2 层的单位面积电容：

$$C_{ox} = \varepsilon_{ox}/d_{ox} \tag{5.8}$$

ε_{ox}和 d_{ox}分别为栅 SiO_2 的介电常数和厚度。

总起来，MOS 结构的平带电压可表示为

$$V_{FB} = \phi_{ms} - Q_0/C_{ox} \tag{5.9}$$

5.1.3 表面势

氧化层下的半导体表面通常简称表面。MOS 器件中所谓的半导体表面指的就是这部分表面。当栅对衬底的外加电压 V_{GB}不等于平带电压 V_{FB}时，半导体将出现表面电荷层，在它之外的半导体内部实际上是电中性的。表面层上的电势降落称为表面势，我们以 ψ_s 表示，并规定电势降落的方向由表面指向体内。因此，若表面电势高于体内，则 ψ_s 为正，反之为负。

利用表面势概念，平衡时表面处的电子浓度和空穴浓度可分别表示为

$$n_s = n_0 e^{\beta\psi_s} \tag{5.10}$$

$$p_s = p_0 e^{-\beta\psi_s} \tag{5.11}$$

n_0 和 p_0 分别是体内中性区的电子和空穴的平衡浓度，$\beta = q/k_B T$。

5.1.4 电势平衡和电荷平衡

对于一般的外加栅压 V_{GB}，半导体表面层将出现电荷并有电势降落，图 5.2 表示外加栅压 V_{GB}时的 p 型衬底 MOS 结构及其电势分布。

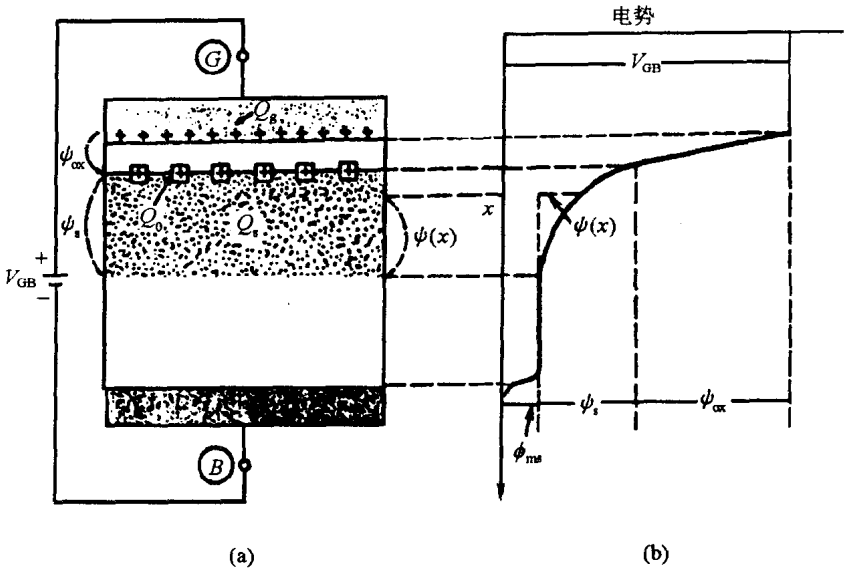


图 5.2 (a) 外加偏压 V_{GB}下的 p 衬底 MOS 结构；(b) 及其电势分布

沿图 5.2(a)所示的电路绕行一周，即得 MOS 结构的电势平衡方程：

$$V_{GB} = \psi_{ox} + \psi_s + \phi_{ms} \tag{5.12}$$

其中，V_{GB}是栅对衬底的偏压；ψ_{ox}是栅氧化层上的电压；ψ_s 是表面势；ϕ_{ms}是衬底对栅的接触电势差，即栅对衬底的等效功函数(费米势)之差。由于 ϕ_{ms}与外加电压无关，所以 V_{GB}改变 ΔV_{GB}时有

$$\Delta V_{GB} = \Delta\psi_{ox} + \Delta\psi_s \tag{5.13}$$

由整体电中性条件，即得 MOS 结构的电荷平衡方程：

$$Q_g + Q_s + Q_0 = 0 \tag{5.14}$$

其中，Q_g 是栅电荷，Q_s 是半导体表面层电荷，Q_0 是有效界面电荷。这里，我们规定上述电荷量都表示电荷面密度(C/cm^2)，即单位表面面积的电荷量。假定 Q_0 不变，则栅电荷 Q_g 改变 ΔQ_g 时有

$$\Delta Q_g + \Delta Q_s = 0 \tag{5.15}$$

另一方面，栅极是 MOS 结构电容器的一个极板(另一个极板是半导体表面层)，其上的电荷量 Q_g 可以通过栅氧化层上的电压降 ψ_{ox}和氧化层电容 C_{ox}表示：

$$Q_g = C_{ox}\psi_{ox} \tag{5.16}$$

5.1.5 半导体表面状态

对于 p 型衬底的 MOS 结构，对应 V_{GB}等于、小于和大于 V_{FB}三种情形，半导体表面将出现平带状态、积累状态和耗尽及反型状态。

(1) 平带状态

$$V_{GB} = V_{FB}, \quad \psi_s = 0, \quad Q_s = 0$$

(2) 积累状态

相对 V_{FB}, V_{GB}有一负的变化量时, Q_g 将产生一个负的变化量。根据方程(5.15),有 $Q_s = -\Delta Q_g > 0$,半导体表面将出现正电荷量 Q_s,这意味着空穴在表面积累;而根据方程(5.13),此时有 $\psi_s < 0$,半导体表面势为负。所以,在积累状态下有

$$V_{GB} < V_{FB}, \quad \psi_s < 0, \ Q_s > 0$$

(3) 耗尽及反型状态

根据类似情形(2)的分析,在表面耗尽或反型状态有

$$V_{GB} > V_{FB}, \quad \psi_s > 0, \ Q_s < 0$$

V_{GB}比 V_{FB}高不多时,正的表面势驱使空穴离开表面,以致表面处的空穴浓度显著低于掺杂浓度;而对电子的吸引作用小到可以忽略不计。所以,这时 Q_s 主要来源于受主离子,可以设想在表面存在一个耗尽层,相当于突变 n^+-p 结 p 型一侧的情形。

V_{GB}增加到使 ψ_s 足够正时,将有大量电子被吸引到表面,以致表面处的电子浓度可能超过空穴浓度。这种状态称为表面反型状态。由于氧化层阻挡了电荷流动,半导体中仍维持平衡状态,即仍有统一的费米能级(E_{FP}),表面处的电子浓度为

$$n_s = n_0 e^{\beta\psi_s} \tag{5.17}$$

利用 $\phi_F = (1/\beta)\ln(N_A/n_i)$和 $n_0 = n_i^2/N_A$,上式可改写为

$$n_s = n_i e^{\beta(\psi_s - \phi_F)} \tag{5.18}$$

或

$$n_s = p_0 e^{\beta(\psi_s - 2\phi_F)} \tag{5.19}$$

p_0 是体内中性区的空穴浓度,这里已假定杂质全部电离,即 $p_0 = N_A$。当 $\psi_s > \phi_F$ 时,表面处的电子浓度超过空穴浓度,表面为反型状态;$\psi_s > 2\phi_F$ 时,表面处的电子浓度超过体内的空穴浓度,通常称为强反型状态。

5.1.6 表面层电荷和表面势的关系

取坐标轴 x,原点选在 SiO_2/Si 界面,方向由界面指向半导体内部,如图 5.2 所示。在 x 处的电子浓度为

$$n(x) = n_0 e^{\beta\psi(x)} \tag{5.20}$$

空穴浓度为

$$p(x) = p_0 e^{-\beta\psi(x)} \tag{5.21}$$

$\psi(x)$是 x 处相对半导体内中性区的电势,即规定中性区内电势为零。表面空间电荷层内的电荷浓度分布为

$$\begin{aligned} \rho(x) &= q[p - n - N_A] \\ &\simeq qN_A[e^{-\beta\psi(x)} - 1 - e^{-2\beta\phi_F}(e^{\beta\psi(x)} - 1)] \end{aligned} \tag{5.22}$$

上式第二步用了中性区内 $p_0 - n_0 - N_A = 0$ 以及 $p_0 \simeq N_A$ 的条件。泊松方程为

$$\frac{d\mathscr{E}(x)}{dx} = \frac{\rho(x)}{\varepsilon_s} \tag{5.23}$$

其中电场强度 $\mathscr{E}(x) = -d\psi(x)/dx$。将方程(5.23)的两边同乘以 $2\mathscr{E}dx$,有

$$d(\mathscr{E}^2) = -\frac{2\rho(x)}{\varepsilon_s} d\psi \tag{5.24}$$

从中性区内任一点($\mathscr{E}=0, \psi=0$)至 x 积分,有

$$\mathscr{E}^2 = -\frac{2}{\varepsilon_s}\int_0^{\psi(x)} \rho(x)\mathrm{d}\psi \tag{5.25}$$

将式(5.22)表示的 $\rho(x)$代入,完成积分,可得

$$\mathscr{E}(x) = \pm \frac{qN_A\sqrt{2}L_D}{\varepsilon_s}\left[\mathrm{e}^{-\beta\psi} + \beta\psi - 1 + \mathrm{e}^{-2\beta\phi_F}(\mathrm{e}^{\beta\psi} - \beta\psi - 1)\right]^{1/2} \tag{5.26}$$

其中

$$L_D \equiv \sqrt{\frac{\varepsilon_s}{\beta q N_A}} \tag{5.27}$$

L_D 称为 p 型半导体的德拜长度。

在方程(5.26)中,令 $\psi(x)=\psi_s$,就得到表面电场强度 $\mathscr{E}_s$。另一方面,根据高斯定理,半导体表面层电荷密度 Q_s 与表面电场 $\mathscr{E}_s$ 的关系为

$$\mathscr{E}_s = -Q_s/\varepsilon_s \tag{5.28}$$

所以,表面层电荷 Q_s 和表面势 ψ_s 的关系为

$$Q_s = \mp qN_A\sqrt{2}L_D\left[\mathrm{e}^{-\beta\psi_s} + \beta\psi_s - 1 + \mathrm{e}^{-2\beta\phi_F}(\mathrm{e}^{\beta\psi_s} - \beta\psi_s - 1)\right]^{1/2} \tag{5.29}$$

方程(5.29)适用于半导体表面的各种状态。"+"号对应 $\psi_s<0$,表示积累;"-"号对应 $\psi_s>0$,表示耗尽或反型。对于每一种实际的表面状态,方程中只有一、二项是主要的,可以大大简化。

5.1.7 表面势和栅压的关系

由前面的讨论可知,对于栅压 V_{GB},超过平带电压 V_{FB}的有效电压($V_{GB}-V_{FB}$)将使半导体表面出现空间电荷层,相应地产生表面势 ψ_s 和表面层电荷 Q_s(栅电荷增加 $-Q_s$),后者在栅氧化层上引起附加电压 $-Q_s/C_{ox}$。换言之,$V_{GB}-V_{FB}$应等于半导体表面势和栅氧化层上附加电压之和,即 $V_{GB}-V_{FB}=\psi_s-Q_s(\psi_s)/C_{ox}$,或者写成下述形式:

$$V_{GB} = V_{FB} + \psi_s - Q_s(\psi_s)/C_{ox} \tag{5.30}$$

其中平带电压 $V_{FB}=\phi_{ms}-Q_0/C_{ox}$。

5.1.8 表面反型状态

在通常的 MOS 器件中,反型层构成导电沟道,是电流导通的原因。所以,表面反型状态对 MOS 器件至关重要。

(1) 反型层电荷

反型时,$\psi_s>\phi_F$,表示表面层电荷 Q_s 的方程(5.29)可简化为

$$Q_s = -qN_A\sqrt{2}L_D\left[\mathrm{e}^{\beta(\psi_s-2\phi_F)} + \beta\psi_s - 1\right]^{1/2} \tag{5.31}$$

在通常的掺杂浓度(例如 $N_A=10^{15}\mathrm{cm}^{-3}$)$F$,$\beta\phi_F=\ln(N_A/n_i)\gg1$,所以 $\beta\psi_s\gg1$,上式可进一步简化为

$$Q_s = -qN_A\sqrt{2}L_D\left[\mathrm{e}^{\beta(\psi_s-2\phi_F)} + \beta\psi_s\right]^{1/2} \tag{5.32}$$

Q_s 包括耗尽层电荷 Q_b 和反型层电荷 Q_i,即

$$Q_s = Q_b + Q_i \tag{5.33}$$

通常,反型载流子(电子)主要分布在紧靠表面的薄层内(其浓度随离开表面的距离指数下

降),其厚度约100Å,比它下面的耗尽层薄得多。所以,一般假定反型层是一个厚度可以忽略的薄层,这一假设称为电荷薄层近似。由于忽略了厚度,反型层上的电势降落也被忽略,即表面势 ψ_s 全部降落在其下的耗尽层上。这样,根据单边突变 p-n 结的理论,耗尽层电荷(C/cm^2)为

$$Q_b = -qN_A W = -\sqrt{2q\varepsilon_s N_A \psi_s} \tag{5.34}$$

式中 $W = \sqrt{2\varepsilon_s \psi_s / qN_A}$ 是表面层(耗尽层)厚度。

式(5.32)和式(5.34)相减,就得到反型层电荷(C/cm^2)为

$$Q_i = -\sqrt{2q\varepsilon_s N_A}\left[\sqrt{\beta^{-1}e^{\beta(\psi_s - 2\phi_F)} + \psi_s} - \sqrt{\psi_s}\right] \tag{5.35}$$

(2) 强反型($\psi_s \geqslant 2\phi_F$)

半导体表面的少数载流子(电子)浓度等于体内的多数载流子(空穴)浓度时,半导体表面开始强反型,这时有

$$\psi_s = 2\phi_F \tag{5.36}$$

$$Q_b = -\sqrt{2q\varepsilon_s N_A (2\phi_F)} \tag{5.37}$$

将式(5.32)和(5.36)代入表示 ψ_s 与 V_{GB} 之间关系的方程(5.30),并令 $V_{GB} = V_{T0}$,可得

$$V_{T0} = V_{FB} + 2\phi_F + \gamma\sqrt{2\phi_F} \tag{5.38}$$

V_{T0} 是半导体表面强反型所需要的最低栅压,称为阈值电压;式中 γ 称为衬偏系数,定义为

$$\gamma \equiv \sqrt{2q\varepsilon_s N_A}/C_{ox} \tag{5.39}$$

V_{GB} 超过 V_{T0} 进一步增加时,反型载流子(电子)电荷迅速增加,具有很高的数值,从而对栅压有很强的屏蔽作用。这时,增加 V_{GB} 基本上只是增加 ψ_{ox},而 ψ_s 变化很小。也就是说,强反型时可认为 ψ_s 具有某一不变的数值($>2\phi_F$),为简单起见,通常设其为 $2\phi_F$,即强反型状态下有

$$\psi_s = 2\phi_F \tag{5.40}$$

相应地,强反型时耗尽层厚度及耗尽层电荷有极大值,分别为

$$X_{dm} = \sqrt{\frac{2\varepsilon_s}{qN_A}(2\phi_F)} \tag{5.41}$$

$$Q_b = -\sqrt{2q\varepsilon_s N_A (2\phi_F)} \tag{5.42}$$

在 ψ_s 固定不变(从而 Q_b 也固定不变)的条件下,过剩栅压($V_{GB} - V_{T0}$)只是形成反型层电荷,即

$$Q_i = -C_{ox}(V_{GB} - V_{T0}) \tag{5.43}$$

5.1.9 电容和电压的关系(*C-V* 特性)

当 V_{GB} 有一微小增量 ΔV_{GB} 时,将有增量电荷 ΔQ_g 流入栅极,同时有等量电荷流出衬底,以维持电荷守恒。这样电荷与电压的变化关系可用一个微分(小信号)电容来表示:

$$C \equiv dQ_g/dV_{GB} \tag{5.44}$$

根据电势平衡方程,有

$$\Delta V_{GB} = \Delta\psi_{ox} + \Delta\psi_s \tag{5.45}$$

即栅压增量的一部分降落在栅氧化层上,使栅氧化层上的电压改变 $\Delta\psi_{ox}$;另一部分降落在半

导体表面空间电荷层上，使表面势改变 $\Delta\psi_s$。用微分增量代替微小增量，就有

$$C = \frac{dQ_g}{d\psi_{ox} + d\psi_s} \tag{5.46}$$

或

$$\frac{1}{C} = \frac{d\psi_{ox}}{dQ_g} + \frac{d\psi_s}{dQ_g} \tag{5.47}$$

如前所述，在有效界面陷阱电荷 Q_0 不随栅压改变的条件下，由电荷平衡方程有 $\Delta Q_s = -\Delta Q_g$，于是式(5.47)可改写成

$$\frac{1}{C} = \frac{1}{dQ_g/d\psi_{ox}} + \frac{1}{-dQ_s/d\psi_s} \tag{5.48}$$

显然，上式右方第一项中的 $dQ_g/d\psi_{ox}$ 为栅氧化层电容 C_{ox}；第二项中的 $-dQ_s/d\psi_s$ 表示半导体表面层电荷 Q_s 随表面势 ψ_s 而变的变化关系，即半导体表面层的电容性质，所以引入半导体表面层电容

$$C_s \equiv -dQ_s/d\psi_s \tag{5.49}$$

因为 Q_s 和 ψ_s 二者的变化方向相反，ψ_s 朝正的方向改变时 Q_s 朝负的方向改变，所以 C_s 之值为正。

由上所述，MOS 二极管的小信号电容等效于氧化层电容 C_{ox} 和半导体表面层电容 C_s 串联，即

$$\frac{1}{C} = \frac{1}{C_{ox}} + \frac{1}{C_s} \tag{5.50}$$

或

$$C = \frac{C_s}{1 + C_s/C_{ox}} \tag{5.51}$$

$C_{ox} = \varepsilon_{ox}/d_{ox}$ 与栅压无关。所以 MOS 二极管的 C-V 特性决定于半导体表面层电容 C_s 随栅压的变化。根据 C_s 的定义和表示 Q_s 的方程(5.29)，可得

$$C_s = \frac{\varepsilon_s}{\sqrt{2}L_D} \frac{\left|1 - e^{-\beta\psi_s} + e^{-2\beta\phi_F}(e^{\beta\psi_s} - 1)\right|}{F(\beta\psi_s)} \tag{5.52}$$

其中

$$F(\beta\psi_s) = \left[e^{-\beta\psi_s} + \beta\psi_s - 1 + e^{-2\beta\phi_F}(e^{\beta\psi_s} - \beta\psi_s - 1)\right]^{1/2} \tag{5.53}$$

图 5.3(a)是 p 型衬底 MOS 二极管 C-V 特性的一条计算曲线。表面积累时，C_s 为

$$C_s \simeq \frac{\varepsilon_s}{\sqrt{2}L_D} e^{-\beta\psi_s/2} \tag{5.54}$$

强积累(ψ_s 很负)状态下 C_s 的数值很大，由式(5.51)可知，$C \simeq C_{ox}$。从物理上讲，是因为这时有大量空穴紧靠氧化层下面，它们构成了栅氧化层电容的一个“极板”(另一个极板是栅极)，结果从 MOS 二极管来看总的电容基本上就是 C_{ox}。当 V_{GB} 增加到使 $\psi_s > 0$ 时，半导体表面进入耗尽和反型状态，$C_s = C_i + C_D$，其中 $C_i = -dQ_i/d\psi_s$ 为反型层电容，$C_D = -dQ_b/d\psi_s$ 为耗尽层电容。在耗尽或反型电荷密度小到可以忽略的弱反型情形下，有

$$C_s \simeq C_D = \sqrt{\frac{\varepsilon_s q N_A}{2\psi_s}} \tag{5.55}$$

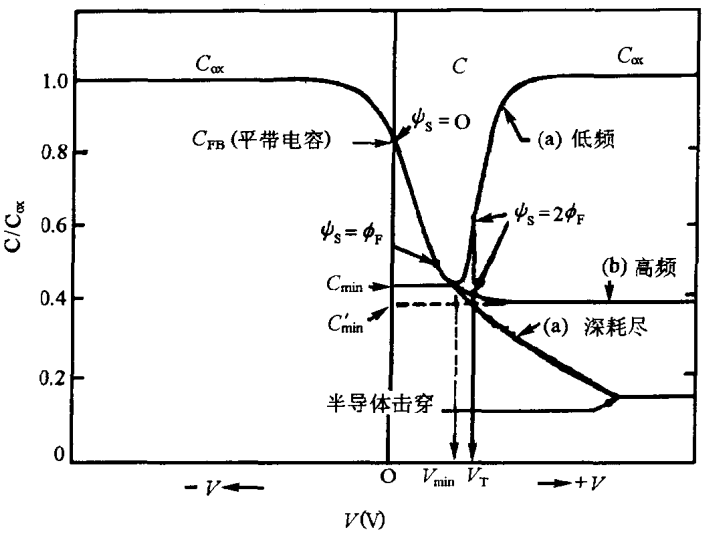


图 5.3　p 衬底 MOS 二极管的 *C-V* 特性

这时，随 V_{GB}(亦即 ψ_s)增加，C_s 下降，故 MOS 二极管电容 C 下降，一直到 ψ_s 接近 $2\phi_F$ 为止。当 V_{GB}增加到使 $\psi_s>2\phi_F$ 时，半导体表面处于强反型状态，

$$C_s \simeq \frac{\varepsilon_s}{\sqrt{2}L_D} e^{\beta(\psi_s-2\phi_F)/2} \tag{5.56}$$

C_s 将很快达到很高的数值，又出现 $C \simeq C_{ox}$的情况，这时大量反型载流子(电子)紧靠氧化层下面构成一个“极板”，所以 MOS 二极管的电容基本上就是 C_{ox}。

C-V 特性是 MOS 二极管的基本特性。通过对 MOS 二极管 *C-V* 特性的测量和分析，可以了解半导体表面状态，了解 SiO_2 层和 SiO_2/Si 界面各种电荷的性质，测定 Si 的许多重要参数(例如掺杂浓度和少数载流子寿命)。实际的 *C-V* 测量中，ΔV_{GB}是一个正弦变化的小信号电压，叠加在变化十分缓慢的栅偏压上。当小信号电压的频率十分低($\sim 10^0$Hz)时，测量的结果和理论计算一致；当小信号电压的频率十分高($\sim 10^3$Hz 或更高)时，反型电荷跟不上 ΔV_{GB} 的变化，测量结果如图 5.3(b)所示。反型层电荷的变化之所以不能足够快地跟上 V_{GB} (ΔV_{GB})的变化，原因是它上面的氧化层和下面的耗尽层使它与外界隔绝了，从而反型层中的电子浓度只能靠热产生-复合过程改变(假定无外部辐照)，而这种过程是十分缓慢的。反之，如果反型层的电荷可以从外部提供或经外部移走(在 MOS 晶体管中，源区和漏区可以提供这种与外界的联系)，则图 5.3(a)所表示的强反型区的 *C-V* 特性可以保持到很高的频率。如果栅偏压和小信号电压 ΔV_{GB}都快速变化，则反型层不能形成，耗尽层厚度将超过强反型时的最大值 X_{dm}，这种状态称为深耗尽状态，其 *C-V* 特性如图 5.3(c)所示。

5.2　MOS 场效应晶体管的基本理论

5.2.1　基本结构和工作原理

图 5.4 表示 n 沟道 MOSFET 的基本结构。两个 n^+ 区便于制作欧姆接触，按使用电位的

高低分别称为源区和漏区，它们之间的距离称为沟道长度(栅长)，通常以 L 表示，图中沿 y 方向；沟道宽度(栅宽)则在与纸面垂直的方向(z 轴方向)，以 Z 表示；与 SiO_2/Si 界面垂直的方向为 x 轴。SiO_2 层把栅和沟道分开，其厚度用 d_{ox}表示。L，Z，d_{ox}是决定 MOSFET 性能的三个重要的设计参数。上述器件的反型层(导电沟道)由电子组成，称为 n 沟 MOSFET，简称 NMOS 或 nMOST；另一类是 p 沟 MOSFET，简称 PMOS 或 pMOST，其导电沟道由空穴组成，衬底是 n 型硅，源区和漏区用 p^+ 区。

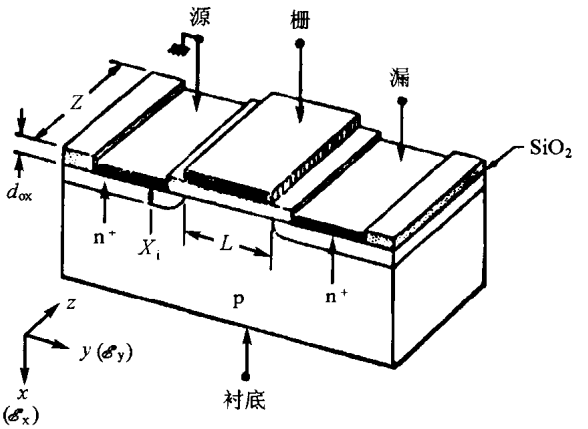


图 5.4　NMOS 的基本结构

MOSFET 除了上述的有 n 沟和 p 沟之分外，还根据栅偏压为零时器件是否导通分为耗尽型(或“常态导通型”)和增强型(或“常态关断型”)。零偏压时关断的称为增强型，这时 NMOS 需要一个正栅压来“增强”沟道，器件才能导通；PMOS 则需要一个负栅压。零偏压时导通的称为耗尽型，这时 NMOS 要求一个负栅压来“耗尽”沟道，器件才能关断；PMOS 则需求一个正栅压。

总起来，MOSFET 有四种类型：n 沟增强型、n 沟耗尽型、p 沟增强型和 p 沟耗尽型。这四类器件的电学符号、输出特性(I_D-V_D 关系)、输入特性(I_D-V_G 关系)表示在图 5.5 中。

使用 MOSFET 时，外加偏压常以源端为参考点，即源端接地。器件正常工作要求源-衬底和漏-衬底这两个 p-n 结反偏，以保证电流沿沟道流动，只有微小的 p-n 结反向漏电流通过衬底。这样，对 NMOS 要求 $V_{BS}\leqslant 0$ 和 $V_{BD}\leqslant 0$。V_{BS}和 V_{BD}分别为衬底相对源和漏的外加电压，由于正常工作时漏对源的偏压 $V_{DS}>0$，所以在 $V_{BS}\leqslant 0$ 的条件下自然满足 $V_{BD}<0$。栅偏压 V_{GS}超过使衬底表面强反型的阈值电压 V_T 时，沟道区表面反型层形成导电沟道，将源和漏沟通，栅压大小控制反型载流子(电子)的数量，从而控制源-漏电流(即漏极电流)I_D，$I_D\sim V_{GS}$ 曲线称为 MOSFET 的输入特性或转移特性。一旦导电沟道形成，加漏偏压 V_{DS}，电子就将通过沟道从源流到漏(相当于电流从漏流到源)，形成漏极电流 I_D。V_{DS}决定沿沟道方向的电场强度，从而决定电子的运动速度，所以 V_{DS}对 I_D 也有控制作用，$I_D\sim V_{DS}$曲线表示 MOSFET 的输出特性，在不同栅压下将得到不同的输出特性曲线。

综合上述，对 n 沟 MOSFET，$V_{GS}-V_T<0$ 时截止，$V_{GS}-V_T>0$ 时处于导通状态，加漏端电压 V_{DS}就会有漏极电流。对 p 沟 MOSFET，工作原理相同，只不过输运电流的反型载流子是空穴，栅压 V_{GS}越负，导电沟道中的空穴浓度越大，所以 $V_{GS}-V_T<0$ 时导通，$V_{GS}-V_T$

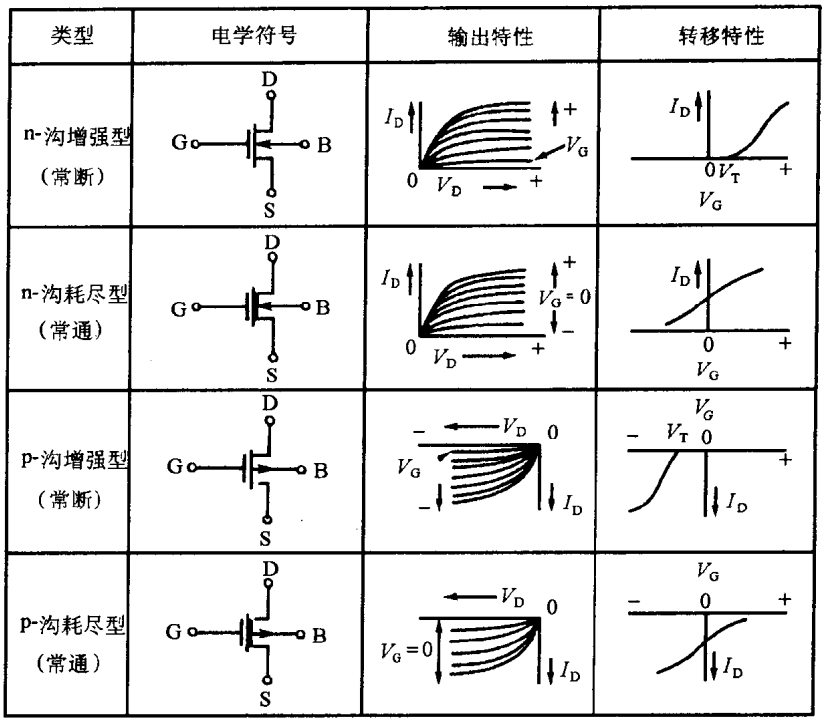


图 5.5　4 种不同类型的 MOSFET

>0 时处于截止状态。另外，正常工作时，p 沟 MOSFET 的漏端应当加负电压，即 $V_{DS}<0$；漏极电流 I_D 的方向与规定的参考方向(从漏到源)相反，即 I_D 为负值。

5.2.2　非平衡状态

对于 MOS 二极管，有外加电压(栅压 V_{GB})时，虽然金属(栅材料)的费米能级 E_{FM}和半导体的费米能级 E_{FP}不再一致($E_{FM}-E_{FP}=-qV_{GB}$)，但因为没有电流流动，半导体从表面到体内仍然具有统一的费米能级，即仍处在平衡状态。而在 MOSFET 中，由于源区和漏区分别与衬底形成 p-n 结，当器件工作时，源区及漏区和表面反型层(沟道)具有相同的导电极性，因此漏区或源区 p-n 结的反向偏置将导致表面沟道与衬底形成的 p-n 结也处于反向偏置状态，并流过一定的反向电流。所以，沟道中载流子(电子)的准费米能级 E_{Fn}与衬底的费米能级 E_{FP}分开，这就是 MOS 器件的非平衡状态。

图 5.6 表示 n 沟 MOSFET 的沟道内任一点 C 处的能带变化，设该点至沟道源端的距离为 y。对于非平衡情形，C 点电子的准费米能级 E_{Fn}相对于衬底的费米能级 E_{FP}下降了 qV_C。为了明确，假定 $V_{DS}=0$，这时整个沟道等电位且与源区电位相同，V_C 就是加在源-衬底结上的反向偏压 V_{BS}。由于表面反型只决定于表面处能级 E_i 和少数载流子(电子)准费米能级 E_{Fn}相交的情形，所以非平衡情形下强反型的表面势不再是 $2\phi_F$，而需要一个更大的表面势，这个表面势应该是 $2\phi_F+V_C(y)$。

当 $V_{DS}\neq 0$ 时，设 C 处相对源的电位为 $V(y)$，则有 $V_C(y)=V(y)-V_{BS}$。需要注意，这时电场沿沟道方向(y 方向)有分量，即电场不再像 MOS 二极管中的情形那样与表面垂直；只

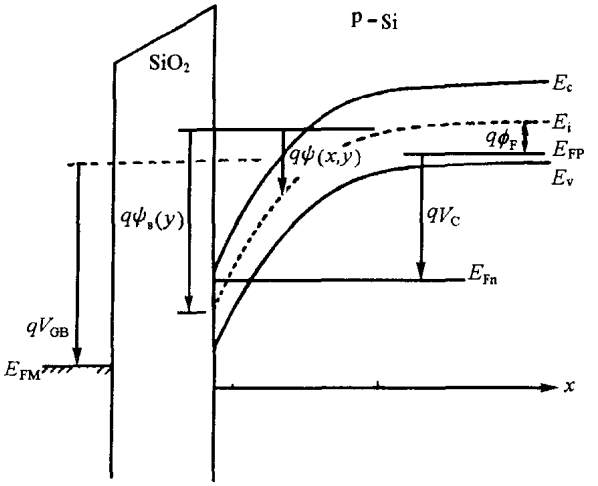


图 5.6　反型时 n 沟 MOSFET 表面附近的能带变化

有假定这一分量比垂直分量小得多，前节中就垂直方向（x 方向）所做的一维分析的结果才可以推广用于 MOSFET。这一假设称为缓变沟道近似（简写为 GCA）。GCA 在沟道长度 L 远大于栅氧化层厚度 d_{ox} 时成立，所以对长沟道器件基本适用，对短沟道器件必须慎重。

在 GCA 下，强反型时有

$$\psi_s(y) = 2\phi_F + V(y) - V_{BS} \tag{5.57}$$

$$Q_b(y) = -\gamma C_{ox}\sqrt{2\phi_F - V_{BS} + V(y)} \tag{5.58}$$

$$Q_i(y) = -C_{ox}\left[V_{GS} - V_{FB} - 2\phi_F - V(y) + \frac{Q_b(y)}{C_{ox}}\right] \tag{5.59}$$

其中 $V_{GS} = V_{GB} + V_{BS}$ 是相对于源的栅压，$V(y)$ 是漏偏压引起的沟道电位分布。

5.2.3　阈值电压

定义为沟道源端的半导体表面开始强反型所需要的栅压。根据定义，它应当由以下三部分组成：(1) 抵消功函数差和有效界面电荷的影响所需栅压，即平带电压 V_{FB}；(2) 产生强反型所需表面势，即 $2\phi_F$；(3) 强反型时栅下表面层电荷 Q_s 在氧化层上产生的附加电压，通常近似为 $-Q_b(2\phi_F)/C_{ox}$。这样，对于 MOSFET，阈值电压的表示式为

$$V_T = V_{FB} + 2\phi_F - Q_b(2\phi_F)/C_{ox} \tag{5.60}$$

需要注意，对 NMOS，$\phi_F = (k_B T/q)\ln(N_A/n_i)$，相应地 $Q_b = -\gamma C_{ox}\sqrt{2\phi_F}$，$\gamma = \sqrt{2\varepsilon_s q N_A}/C_{ox}$；对 PMOS，$\phi_F = -\dfrac{k_B T}{q}\ln(N_D/n_i)$，相应地 $Q_b = \gamma_p C_{ox}\sqrt{(-2\phi_F)}$，$\gamma_p = \sqrt{2\varepsilon_s q N_D}/C_{ox}$。$N_A$，$N_D$ 是半导体衬底的掺杂浓度。

在 MOS 集成电路的设计和生产中，阈值电压的控制十分重要，首先必须能够按照电路要求可靠地生产出增强型或耗尽型器件。大多数应用中需要的是增强型器件，这时对 NMOS 要求 $V_T > 0$，对 PMOS 要求 $V_T < 0$。上述要求对 PMOS 容易达到，对 NMOS 却很困难，原因是 $V_{FB} = \phi_{ms} - Q_0/C_{ox}$ 中的 Q_0 总是正的，即 $-Q_0/C_{ox}$ 总是负的，结果使 p 沟器件和 n 沟器件的 V_{FB} 一般都是负的。对 PMOS，V_T 表示式中的另外两项也是负的，因此生产增强型没有困难；

对 NMOS，另外两项之和必须大于 V_{FB}，从而要求衬底掺杂浓度较高，这会导致大的衬底电容和低的击穿电压，是十分不理想的。

为了有效地调整阈值电压，常常使用离子浅注入的方法，即通过栅氧化层把杂质注入到沟道表面的薄层内，其作用相当于有效界面电荷，所以阈值电压的改变可以从下面的公式估算：

$$\Delta V_T \simeq \pm qN_I/C_{ox} \tag{5.61}$$

N_I 是注入剂量(离子个数/cm^2)。注入杂质为 p 型时，上式取正号；n 型时取负号。

施加反向衬底偏压也能调整阈值电压。对于 n 沟器件，这时沟道源端在强反型时的耗尽层电荷为

$$Q_b = -\gamma C_{ox}\sqrt{2\phi_F - V_{BS}} \tag{5.62}$$

V_{BS}表示衬底相对于源的外加电压，n 沟器件的衬底是 p 型，$V_{BS}<0$ 时为反向衬底偏压(p 沟器件的反向衬底偏压 $V_{BS}>0$)。相应的阈值电压为

$$V_T' = V_{FB} + 2\phi_F + \gamma\sqrt{2\phi_F - V_{BS}} \tag{5.63}$$

所以衬底偏压引起的阈值电压增量为

$$\Delta V_T = \gamma\left[\sqrt{2\phi_F - V_{BS}} - \sqrt{2\phi_F}\right] \tag{5.64}$$

$\gamma = \sqrt{2\varepsilon_s qN_A}/C_{ox}$反映衬底偏压对阈值电压影响的强弱程度，故称作衬偏系数。对于 p 沟器件，衬底偏压引起的阈值电压改变为

$$\Delta V_T = -\gamma_p\left[\sqrt{|2\phi_F| + V_{BS}} - \sqrt{|2\phi_F|}\right] \tag{5.65}$$

$\gamma_p = \sqrt{2\varepsilon_s qN_D}/C_{ox}$是 p 沟器件(n 型衬底)的衬偏系数。

氧化层厚度对阈值电压也有影响。氧化层厚度增加时，栅压对半导体表面的控制作用减弱，为了使表面形成导电沟道，需要更大的栅压，即阈值电压增加。这一点对 MOS 器件以外区域的半导体表面十分重要，这些区域称为场区。图 5.7 表示一个未经金属化的 n 沟 MOSFET 的透视图，其顶层是掺磷二氧化硅(磷硅玻璃)，用作多晶硅栅与栅金属引线之间的绝缘层。由该图可见，场区氧化层比栅氧化层厚得多，为 1μm 左右；场区的阈值电压可高达十几伏，比栅压大一个数量级。因此，若将 5V 电压同时加到栅氧化层上和场氧化层上，栅下形成了导电沟道，而场氧化层下的半导体表面仍保持耗尽状态。这样，场氧化层适用于 MOS 器件之间的隔离。在集成电路中，由于内部引线要在氧化层上面走过，这些引线有可能承受很高的电压，使场区的半导体表面仍有可能反型，产生寄生沟道，导致电路不能正常工作。为了防止寄生沟道，场区还必须进行高浓度掺杂(掺与衬底同类型的杂质)，使表面不容易反型，从而将沟道隔断。

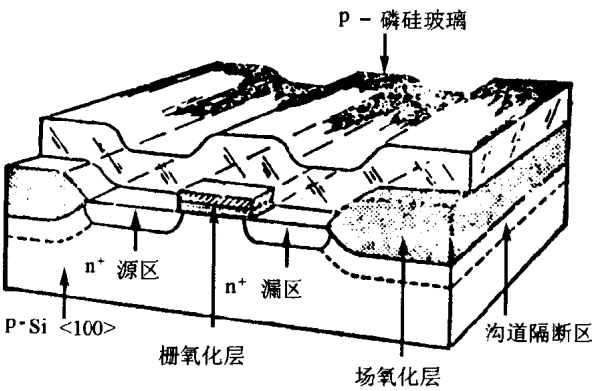


图 5.7　n 沟 MOSFET 透视图

5.2.4 电流基本特性

MOSFET 是一种电压控制器件，漏极电流 I_D 受栅压 V_{GS}和漏压 V_{DS}控制。对器件导通状态的分析一般采用强反型近似，这时 I_D 主要来自反型载流子的漂移作用。对接近截止的状态，反型载流子很少，采用弱反型近似，并认为 I_D 主要来自反型载流子的扩散作用，相应的电流特性称为亚阈区特性。作为对电流特性的基本分析，我们采用缓变沟道近似。此外，仍以 n 沟 MOSFET 为例。

1. 电流方程

MOSFET 的电流方程表示漏极电流同栅压及漏压的关系。这里讨论器件的导通状态。设漏压 V_{DS}很小，沟道内的电场强度远低于速度饱和电场，于是漏极电流可表示为

$$I_D = \mu_n Z(-Q_i)\frac{dV}{dy} \tag{5.66}$$

其中，Z 为沟道宽度，$Z(-Q_i)$是单位沟道长度内的电荷数，即反型电荷沿沟道方向(y 方向)的分布；μ_n 在这里表示反型载流子(电子)的表面迁移率，比体内的电子迁移率小得多，原因来自栅压产生的垂直电场把电子拉向表面以及表面的附加散射，下面的讨论中假设 μ_n 为常数，其大小近似为体内迁移率的 1/2；$\mathscr{E}(y) = -dV(y)/dy$ 是沿沟道方向的电场。规定 I_D 的参考方向是从漏通过沟道到源，这一方向与我们选择的 y 方向相反。

由于电流连续，I_D 与位置 y 无关。将方程(5.66)从 $y=0$ 至 $y=L$ 积分，有

$$I_D = \frac{Z}{L}\mu_n\int_0^{V_{DS}}(-Q_i)dV \tag{5.67}$$

$Q_i(y)$已由方程(5.59)给出，将它代入上式并完成积分后，就可以得到漏极电流 I_D 的一般方程。可以预期这一方程是比较复杂的，原因是 $Q_i(y)$的表示式中包含 $Q_b(y)/C_{ox}$而出现平方根项 $\gamma\sqrt{2\phi_F - V_{BS} + V(y)}$。所以，如果对 Q_b/C_{ox}作适当近似，将会得到比较简单的电流方程。通常的做法是将 $Q_b/C_{ox} = -\gamma\sqrt{2\phi_F - V_{BS} + V(y)}$在 $V=0$ 附近展开成级数并只取前二项，即

$$-\frac{Q_b(y)}{C_{ox}} = \gamma\sqrt{2\phi_F - V_{BS}} + \alpha V(y) \tag{5.68}$$

其中

$$\alpha = C_D/C_{ox}, \quad C_D = \frac{1}{2}\sqrt{\frac{2\varepsilon_s q N_A}{2\phi_F - V_{BS}}} \tag{5.69}$$

C_D 是表面势 $\psi_s = 2\phi_F - V_{BS}$时的沟道耗尽层电容。于是，将式(5.68)代入方程(5.59)，我们将得到

$$Q_i(y) = -C_{ox}[(V_{GS} - V_T) - (1+\alpha)V(y)] \tag{5.70}$$

$V(y)$是 $V_{DS}\neq 0$ 时沟道内 y 处相对源端的电位，V_T 是器件的阈值电压，即

$$V_T = V_{FB} + 2\phi_F + \gamma\sqrt{2\phi_F - V_{BS}} \tag{5.71}$$

在方程(5.70)中令 $\alpha=0$，就得到使用最广泛的电压控制模型：

$$Q_i(y) = -C_{ox}[V_{GS} - V_T - V(y)] \tag{5.72}$$

在以后的讨论中，我们将采用这一模型。

2. 线性区与饱和区

将式(5.72)代入方程(5.67)，可以得到非饱和区的电流方程(Sah 方程)

$$I_D = \frac{Z}{2L}\mu_n C_{ox}[2(V_{GS} - V_T)V_{DS} - V_{DS}^2] \tag{5.73}$$

方程中方括号前的因子常称作导电因子，以 β 表示，即 $\beta = \frac{Z}{2L}\mu_n C_{ox}$。

线性区 V_{DS}很小时，方程(5.73)简化为

$$I_D = \frac{Z}{L}\mu_n C_{ox}(V_{GS} - V_T)V_{DS} \tag{5.74}$$

这时，从源到漏形成比较均匀的导电沟道，整个沟道相当于一个阻值正比于$(V_{GS} - V_T)^{-1}$的电阻，栅压 V_{GS}越大，阻值越小。通过沟道的电流随 V_{DS}线性增加，故这一工作区称为线性区。

在线性区，器件的跨导为

$$g_m \equiv \left.\frac{\partial I_D}{\partial V_{GS}}\right|_{V_{DS}} = \frac{Z}{L}\mu_n C_{ox} V_{DS} \tag{5.75}$$

沟道电导为

$$g_D \equiv \left.\frac{\partial I_D}{\partial V_{DS}}\right|_{V_{GS}} = \frac{Z}{L}\mu_n C_{ox}(V_{GS} - V_T) \tag{5.76}$$

饱和区 当 V_{DS}增加到 $V_{Dsat} = V_{GS} - V_T$ 时，沟道漏端的反型电荷 $Q_i = 0$，即漏端已经不能形成反型层，沟道在漏端夹断。随着 V_{DS}继续不断增加，夹断点从漏端逐渐移向源端，所增加的电压$(V_{DS} - V_{Dsat})$降落在夹断点和漏端之间的耗尽区(夹断区)上，在这里形成强电场区。沟道夹断以后，源端和夹断点之间仍存在导电沟道，载流子漂移到夹断点就被这里的强电场拉到漏极，这十分类似双极晶体管中载流子通过反偏的集电结耗尽区的情况。由于夹断以后导电沟道上的电压基本上维持在 V_{Dsat}不变，漏极电流基本不变而达到饱和值 I_{Dsat}：

$$I_{Dsat} = \frac{Z}{2L}\mu_n C_{ox}(V_{GS} - V_T)^2 \tag{5.77}$$

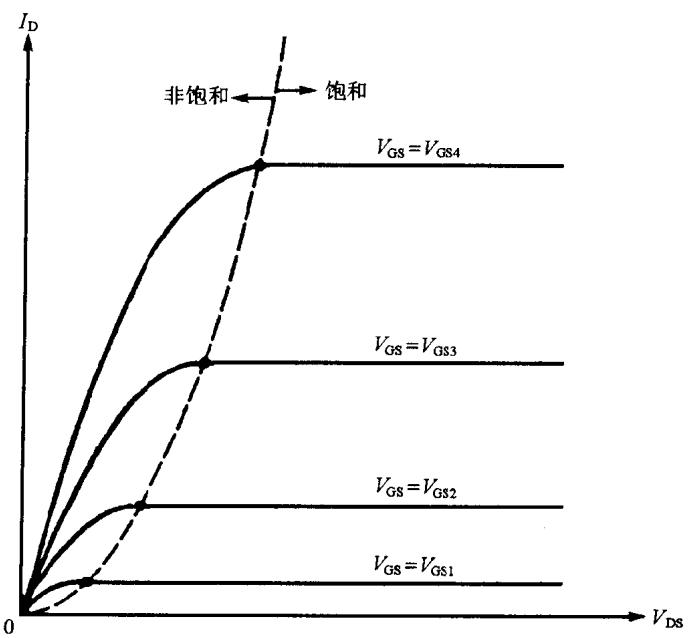


图 5.8 n沟 MOSFET 的 I_D-V_{DS}特性

$V_{DS} \geqslant V_{Dsat}$的区域称为饱和区。在饱和区有

$$g_m = \frac{Z}{L}\mu_n C_{ox}(V_{GS} - V_T) \tag{5.78}$$

在理想情形下，由于 I_D 在这时不再是 V_{DS}的函数，所以饱和区的沟道电导 $g_D = 0$；实际上，由于有效沟道长度缩短，饱和区有非零的沟道电导。

图 5.8 中的 I_D-V_{DS}曲线是根据方程(5.73)和(5.77)画出的。对给定的 V_{GS}，V_{DS}很小时 $I_D \propto V_{DS}$(线性区)；随 V_{DS}增加，I_D-V_{DS}曲线逐渐平坦而趋于饱和值。图中虚线表示 I_D 达到最大值 I_{Dsat}时 V_{Dsat}的轨迹。在饱和区，$I_D \propto (V_{DS}-V^T)^2$，对不同的栅压 V_{GS}，I_D-V_{GS}曲线的间距随$(V_{GS}-V_T)^2$ 变化。

3. 亚阈值区

栅偏压低于阈值以致沟道区表面是弱反型时，MOSFET 仍有很小的漏极电流，这就是亚阈值电流，器件的工作状态也被说成是在亚阈值区。亚阈值区描述 MOSFET 是如何导通和截止的，所以在器件的开关或数字电路应用中亚阈值区特性是十分重要的。

在亚阈值区，对漏极电流起决定作用的是扩散而不是漂移。事实上，MOSFET 在弱反型时和双极晶体管有十分相似之处，其中沟道区起基区作用，n^+ 源区和 n^+ 漏区分别起发射区和集电区的作用，外加“集电极-发射极”电压(漏电压 V_{DS})主要降落在反偏结(即漏-衬底结)上。所以，亚阈值电流可以像在均匀基区的双极晶体管中求注入基区的少数载流子(电子)电流那样得到，

$$J_n = -qD_n \frac{n(y_S) - n(y_D)}{L - y_S - y_D} \tag{5.79}$$

其中 y_S 和 y_D 分别为源结和漏结的水平耗尽区宽度(参阅图 5.19)；$L - y_S - y_D$ 为有效沟道长度；$n(y_S)$和 $n(y_D)$分别为有效沟道源端和漏端的电子浓度，

$$n(y_S) = N_D e^{\beta V_S^0} \tag{5.80}$$

$$n(y_D) = N_D e^{\beta(V_S^0 - V_{DS})} \tag{5.81}$$

$\beta = q/k_B T$，N_D 是 n^+ 区的掺杂浓度。V_S^0 是沟道源端相对源接触($y = 0$)的表面势，亦即源结(n^+-p 结)水平耗尽区上的电压为 $-V_S^0$。V_S^0 和常规意义下(相对衬底内部而言)的表面势 ψ_s 之间的关系显然为 $\psi_s = V_S^0 + V_{bi}$，其中 V_{bi}是 n^+-p 结的内建电压。

将式(5.80)和(5.81)代入式(5.79)，有

$$J_n = -\frac{qD_n N_D e^{\beta V_S^0}[1 - e^{-\beta V_{DS}}]}{L - y_S - y_D} \tag{5.82}$$

从漏端进入的方向为漏极电流方向，故亚阈值电流可表示为

$$I_D = \frac{Z\delta q D_n N_D e^{\beta V_S^0}[1 - e^{-\beta V_{DS}}]}{L - y_S - y_D} \tag{5.83}$$

在上式中，δ 是有效沟道厚度，定义为表面垂直方向(x 方向)上电势减小 $k_B T/q$ 的距离。设 δ 上的电场强度是常数且等于表面电场 $\mathscr{E}_s$，则有

$$\delta \cdot \mathscr{E}_s = k_B T/q = 1/\beta \tag{5.84}$$

而

$$\mathscr{E}_s = -\mathrm{d}\psi_s/\mathrm{d}x = -\frac{\mathrm{d}\psi_s}{\mathrm{d}Q_s} \cdot \frac{\mathrm{d}Q_s}{\mathrm{d}x} \simeq -\frac{1}{\mathrm{d}Q_B/\mathrm{d}\psi_s} \cdot \frac{\mathrm{d}Q_b}{\mathrm{d}x} = \frac{qN_A}{C_D} \tag{5.85}$$

$C_D \equiv -\mathrm{d}Q_b/\mathrm{d}\psi_s$ 是表面耗尽层电容，利用 $Q_b = -\sqrt{2\varepsilon_s q N_A \psi_s}$，得到

$$C_D = \sqrt{\frac{\varepsilon_s q N_A}{2\psi_s}} \tag{2.86}$$

联合式(5.84)和(5.85)，我们得到

$$\delta = \frac{C_D(\psi_s)}{\beta q N_A} \tag{5.87}$$

栅压低于阈值时，反型电荷很少，$Q_s \simeq Q_b$。将 $Q_b(\psi_s)$ 在 $\psi_s = 2\phi_F$ 附近展开，代入下述栅压和表面势之间关系的表示式

$$V_{GS} \simeq V_{FB} + \psi_s - Q_b(\psi_s)/C_{ox}$$

可以得到

$$\psi_s - 2\phi_F = V_{GT}/\eta \tag{5.88}$$

其中 $V_{GT} = V_{GS} - V_{T0}$，V_{T0} 是无衬底偏压时 MOSFET 的阈值电压；η 称为亚阈值理想因子，

$$\eta = 1 + C_D/C_{ox} \tag{5.89}$$

由式(5.88)和 $\psi_s = V_S^0 + V_{bi}$，得到

$$V_S^0 = -V_{bi} + 2\phi_F + V_{GT}/\eta \tag{5.90}$$

将式(5.87)和(5.90)代入式(5.83)，并加以整理，最终可将亚阈值电流表示为

$$I_D = \frac{Z\mu_n(k_B T/q)^2 C_D(\psi_s) e^{qV_{GT}/\eta k_B T}[1 - e^{qV_{DS}/k_B T}]}{L - y_S - y_D} \tag{5.91}$$

其中

$$y_S = \sqrt{\frac{2\varepsilon_s}{qN_A}(-V_S^0)} \tag{5.92}$$

$$y_D = \sqrt{\frac{2\varepsilon_s}{qN_A}(-V_S^0 + V_{DS})} \tag{5.93}$$

由于栅压引起的纵向电场 $\mathscr{E}_t$ 强烈影响表面势分布，y_S 和 y_D 分别小于源结和漏结的垂直耗尽区宽度 W_S 和 W_D。对于长沟道器件，$L \gg y_S + y_D$，可以把方程(5.91)的分母中的 y_S 和 y_D 略去。这时由该式可以看到，$V_{DS} > 3k_B T/q$ 时亚阈值电流与 V_{DS} 无关。

为了表征亚阈值电流随栅压的变化，引进下面的参数：

$$S \equiv \frac{\mathrm{d}V_{GS}}{\mathrm{d}(\lg I_D)} = \frac{\mathrm{d}V_{GS}}{\mathrm{d}(\ln I_D)}\ln 10 \tag{5.94}$$

S 称为亚阈值斜率，它表示 I_D 改变一个数量级所需要的栅压摆幅。S 越小，器件导通和截止之间的转换就越容易，说明亚阈值区特性越好。对于长沟道 MOSFET，在表示亚阈值电流的公式(5.91)中忽略 y_S 和 y_D 的作用，由式(5.94)可得

$$S = \frac{k_B T}{q}(\ln 10)\left[1 + \frac{C_D}{C_{ox}}\right] \tag{5.95}$$

下面我们证明关于长沟道 MOSFET 亚阈值斜率的一个有用的表示式：

$$S = \frac{k_B T}{q}(\ln 10)\frac{\mathrm{d}V_G}{\mathrm{d}\psi_s} \tag{5.96}$$

为此，利用 $V_S^0 = \psi_s - V_{bi}$ 和 $\delta = \frac{C_D(\psi_s)}{\beta q N_A}$，以及 $\beta V_{bi} = \ln\left(\frac{N_A N_D}{n_i^2}\right)$，把表示 I_D 的方程(略去分母中的

y_S 和 y_D)改写为 ψ_s 的函数,如下式:

$$I_D = \mu_n\left(\frac{Z}{L}\right)\left(\frac{n_i}{N_A}\right)^2 \frac{C_D(\psi_s)}{\beta^2} - e^{\beta\psi_s}[1 - e^{-\beta V_{DS}}]$$

由方程(5.97)可推出

$$\frac{d(\ln I_D)}{dV_G} = \left(\beta - \frac{1}{2\psi_s}\right)\frac{d\psi_s}{dV_G} \simeq \beta\frac{d\psi_s}{dV_G} \tag{5.98}$$

进一步,根据 S 的定义式(5.94),就可证明公式(5.96)。

我们可以利用沟道区的电容分压电路来得到 $dV_G/d\psi_s$。图 5.9 是忽略界面陷阱作用时的电容分压模型。由图不难看出,有

$$C_{ox}\,d(V_G - \psi_s) = C_D\,d\psi_s$$

即

$$\frac{dV_G}{d\psi_s} = 1 + \frac{C_D}{C_{ox}}$$

所以有

$$S = \frac{k_B T}{q}(\ln 10)\left[1 + \frac{C_D}{C_{ox}}\right]$$

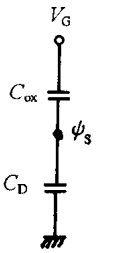


图 5.9 沟道区的电容分压模型

与前面计算得到结果(5.95)式一致。由 $C_D = \left(\frac{\varepsilon_s q N_A}{2\psi_s}\right)^{1/2}$ 可知,衬底掺杂浓度 N_A 越低,C_D/C_{ox} 越小,S 也就越小,器件导通和截止之间越容易转换。

界面陷阱浓度很高时,还必须考虑界面陷阱电容 C_{it}的作用,这是因为界面陷阱能级随表面势变化相对半导体的费米能级移动,因此在表面附近同半导体交换电荷,具有电容的作用。C_{it}和 C_D 并联,在亚阈区斜率的表达式(5.95)中以($C_D + C_{it}$)替代 C_D,得到

$$S(\text{有界面陷阱}) = S(\text{无界面陷阱}) \times \frac{1 + (C_D + C_{it})/C_{ox}}{1 + C_D/C_{ox}} \tag{5.99}$$

4. *表面迁移率的修正*

前面已经指出,反型层中载流子的迁移率降低到仅为其体内迁移率的一半左右。迁移率降低的原因是载流子在沟道中运动时会受到栅压 V_{GS}引起的纵向电场 $\mathscr{E}_x$ 和漏压 V_{DS}引起的横向电场 $\mathscr{E}_y$ 的影响。$\mathscr{E}_x$ 把载流子吸引到表面,从而加剧了表面散射而引起迁移率降低;$\mathscr{E}_y$ 则使载流子的迁移率和横向电场有关,甚至会达到速度饱和。这里,我们只考虑栅压电场 $\mathscr{E}_x$ 的影响,漏压电场 $\mathscr{E}_y$ 的作用则在讨论速度饱和效应时考虑。

一般说来,$\mathscr{E}_x$ 沿沟道变化,所以表面迁移率 μ_n 也是沿沟道变化的,它可以用下述经验公式表示:

$$\mu_n = \frac{\mu_0}{1 + a_\theta \bar{\mathscr{E}}_x} \tag{5.100}$$

μ_0 是低栅压电场下的表面迁移率,$a_\theta \simeq 0.025\mu m/V$,$\bar{\mathscr{E}}_x$ 是沟道上纵向场强的平均值:

$$\bar{\mathscr{E}}_x = -\frac{Q_n + 2Q_B}{2\varepsilon_s} \tag{5.101}$$

现在,表示漏极电流的方程(5.66)可以修改为

$$I_D = \frac{\mu_0}{1 - \frac{a_\theta}{2\varepsilon_s}(Q_n + 2Q_B)} Z(-Q_n)\frac{dV}{dy} \tag{5.102}$$

引入修正的表面迁移率 μ_{eff}:

$$\mu_{eff} = \frac{\mu_0}{\frac{1}{L}\int_0^L \left[1 - \frac{a_\theta}{2\varepsilon_s}(Q_n + 2Q_B)\right]dy} \tag{5.103}$$

则有

$$I_D = \frac{Z}{L}\mu_{eff}\int_0^{V_{DS}}(-Q_n)dy \tag{5.104}$$

这一结果表明,与常数表面迁移率 μ_n 的情形相比,只是用修正的迁移率(有效迁移率)μ_{eff}代替了 μ_n。

为了得到 μ_{eff}的近似表达式,我们假定沟道电位 $V(y)$随 y 线性变化,

$$V(y) = \frac{V_{DS}}{L}y \tag{5.105}$$

则 μ_{eff}可表示为

$$\mu_{eff} = \frac{\mu_0}{\frac{1}{V_{DS}}\int_0^{V_{DS}} \left[1 - \frac{a_\theta}{2\varepsilon_s}(Q_n + 2Q_B)\right]dy} \tag{5.106}$$

对于电压控制模型有

$$\mu_{eff} = \frac{\mu_0}{1 + \theta f_\mu} \tag{5.107}$$

其中

$$\theta = \frac{a_\theta}{2\varepsilon_s}C_{ox} \tag{5.108}$$

$$f_\mu = (V_{GS} - V_T) + 2\gamma\sqrt{2\phi_F - V_{BS}} - V_{DS}/2 \tag{5.109}$$

有时,对 f_μ 作进一步近似,这就是略去 V_{DS},并用 $\theta_B V_{BS}$代替 $2\gamma\sqrt{2\phi_F - V_{BS}}$。于是,

$$\mu_{eff} = \frac{\mu_o}{1 + \theta(V_{GS} - V_T) + \theta_B V_{SB}} \tag{5.110}$$

θ_B 的数值一般为每伏百分之几。这里的 μ_{eff}只适用于载流子漂移速度(或$\mathscr{E}_y$)较小的情形。

5. *温度效应*

温度影响器件的参数和性能,特别是影响迁移率、阈值电压和亚阈值区特性。沟道载流子的迁移率随温度增加而下降,常用的近似公式为

$$\mu(T) = \mu(T_0)(T/T_0)^{-m} \tag{5.111}$$

T 为绝对温度,$T_0 = 300K$, m 的数值在 1.5~2.0。

阈值电压用下式表示

$$V_T = V_{FB} + 2\phi_F - Q_b(2\phi_F)/C_{ox}$$

V_{FB}基本上与温度无关。上式对温度的微商为

$$\frac{dV_T}{dT} \simeq \left[2 - \frac{Q_b}{2C_{ox}\phi_F}\right]\frac{d\phi_F}{dT} \tag{5.112}$$

对 p 型衬底的器件，$\phi_F=(k_BT/q)\ln(N_A/n_i)$，其中本征载流子浓度 n_i 的温度关系为

$$n_i \sim T^{3/2}\, e^{-E_g/2k_BT}$$

故可得到

$$\frac{d\phi_F}{dT} \simeq \frac{1}{T}[\phi_F - E_g/2q] \tag{5.113}$$

写出上式时用到了一般温度下 $E_g/2k_BT \gg 3/2$ 的条件。

对于 n 型衬底的器件，$\phi_F=-(k_BT/q)\ln(N_D/n_i)$，通过类似于 p 型衬底情形下的推导过程，可得

$$\frac{d\phi_F}{dT} \simeq \frac{1}{T}[-|\phi_F| + E_g/2q] \tag{5.114}$$

综合上述，MOSFET 的阈值电压随温度的变化可统一表示为

$$\frac{dV_T}{dT} \simeq \pm \frac{1}{T}\left[2 - \frac{Q_b}{2C_{ox}\phi_F}\right]\left[\frac{E_g(T=0)}{2q} - |\phi_F|\right] \tag{5.115}$$

“－”号用于 n 沟器件，“＋”号用于 p 沟器件。

通过迁移率和阈值电压的温度关系，可以分析 MOSFET 的性能同温度的关系。随着温度下降，器件特性将明显改善，最重要的改善是亚阈值斜率 S 减小，这主要来自其表示式中的因子 k_BT/q。

5.2.5 瞬态特性

前面讨论的电流基本特性为 MOSFET 的稳态或直流特性，下面将讨论 MOSFET 的瞬态特性。

当信号电压加到直流偏压上时，或者说各个端电压发生变化时，MOSFET 的栅电荷和耗尽层电荷将发生变化，漏极电流也将发生变化。前者是电容效应，后者是电流效应。MOSFET 的瞬态特性由器件的电容效应引起。当外加电压变化的幅度很小时，分析瞬态特性的模型为小信号模型；当电压变化的幅度较大时，相应的模型为大信号模型。我们的讨论限于小信号情形，而且只考虑器件的“本征”部分。

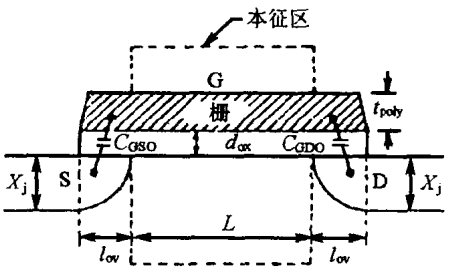


图 5.10 MOSFET 的本征区

MOSFET 的本征部分指源区和漏区之间的部分，包括栅极、栅氧化层及其下面的耗尽层，如图 5.10 中的虚线框所示。虚线框以外的部分为非本征部分，引起寄生效应，例如由于栅和两个 n^+ 区之间的重叠是不可避免的，这就产生寄生的覆盖电容。注意到栅和 n^+ 区都是高浓度掺杂，覆盖电容可用电介质为栅氧化层的平板电容器 C_{GSO}和 C_{GDO}来模拟，下标“O”代表非本征的。如果重叠长度是 l_{ov}，栅宽是 Z，则忽略边缘效应(即由于电场线不是垂直于表面，而是沿着重叠区之外的边缘而产生的效应)以后，我们有

$$C_{GSO} = C_{GDO} = C_{ox} Z l_{ov}$$

如果 l_{ov}不很小，则由于边缘电场而引起的电容可以使总电容值增加一个不小的百分数。这就牵涉到边缘电容的计算问题，并常常用经验公式估算。又如，由于源结和漏结都是 p-n 结，这就存

在 p-n 结电容 C_{BSO}和 C_{BDO};而对 CMOS(互补晶体管)的阱内器件,还必须考虑第三个 p-n 结电容,即阱和公共衬底之间的 p-n 结而形成的电容 $C_{B'B}$(B 对应晶体管本体,B′对应公用衬底)。器件的功能与本征部分有关,非本征部分引起寄生效应,应尽可能降低。

1. 本征电容

栅-源电容 C_{GS} 定义为

$$C_{GS} \equiv -\left.\frac{\partial Q_G}{\partial V_S}\right|_{V_G, V_D, V_B} \tag{5.114}$$

可理解为,源电压增加 ΔV_S 时,栅氧化层上的电压减小 ΔV_S,从而栅上的总电荷量减少 ΔQ_G。

源-衬底电容 C_{BS} 定义为

$$C_{BS} \equiv -\left.\frac{\partial Q_B}{\partial V_S}\right|_{V_G, V_D, V_B} \tag{5.115}$$

可理解为,源电压增加 ΔV_S 时,耗尽层厚度增加,其总电荷量 Q_B 更负,或者说减少 ΔQ_B。

栅-漏电容 C_{GD} 定义为

$$C_{GD} \equiv -\left.\frac{\partial Q_G}{\partial V_D}\right|_{V_G, V_S, V_B} \tag{5.116}$$

漏-衬底电容 C_{BD} 定义为

$$C_{BD} \equiv -\left.\frac{\partial Q_B}{\partial V_D}\right|_{V_G, V_S, V_B} \tag{5.117}$$

增加 V_D 和增加 V_S 的效应相同,因此 C_{GD}和 C_{GS}、C_{BD}和 C_{BS}有类似的物理意义。

栅-衬电容 C_{GB} 定义为

$$C_{GB} \equiv -\left.\frac{\partial Q_G}{\partial V_B}\right|_{V_G, V_D, V_S} \tag{5.118}$$

可理解为,衬底偏压增加 ΔV_B 时,引起附加的正电荷流入栅下耗尽层,这些电荷部分地由栅电荷量 Q_G 的减少来平衡。

上面定义的 5 个电容依赖于各个端电压的数值,求出总电荷量 Q_G 和 Q_B 同它们的关系就可以计算这些电容。

栅下耗尽层电荷总量 Q_B 可表示为

$$Q_B = \int_0^L Q_b(Z\mathrm{d}y) = Z\int_0^L Q_b\mathrm{d}y \tag{5.119}$$

Q_b 是单位面积沟道区耗尽层的电荷量。利用 $I_D = \mu_n Z(-Q_i)(\mathrm{d}V/\mathrm{d}y)$,可将上式改写成

$$Q_B = -\frac{\mu_n Z^2}{I_D}\int_0^{V_{DS}} Q_b Q_i \mathrm{d}V \tag{5.120}$$

同理,反型层电荷总量 Q_I 可表示成

$$Q_I = -\frac{\mu_n Z^2}{I_D}\int_0^{V_{DS}} Q_i^2 \mathrm{d}V \tag{5.121}$$

栅电荷总量 Q_G 通过下述关系得到

$$Q_G + Q_I + Q_B + Q_O = 0 \tag{5.122}$$

其中有效界面电荷总量 $Q_O = ZLQ_0$,Q_0 是单位面积有效界面电荷,ZL 是沟道区面积。

2. 跨导和沟道电导

(1) 跨导 g_m 定义为

$$g_m \equiv \frac{\partial I_D}{\partial V_{GS}}\bigg|_{V_{DS},\,V_{BS}} \tag{5.123}$$

用电压控制模型,有

$$g_m = \begin{cases} \dfrac{Z}{L}\mu_n C_{ox} V_{DS} & \text{(非饱和区)} \\ \dfrac{Z}{L}\mu_n C_{ox}(V_{GS} - V_T) & \text{(饱和区)} \end{cases} \tag{5.124}$$

(2) 衬底跨导 g_{mB} 定义为

$$g_{mB} \equiv \frac{\partial I_D}{\partial V_{BS}}\bigg|_{V_{GS},\,V_{DS}} \tag{5.125}$$

用电压控制模型,容易证明

$$g_{mB} = g_m \cdot \gamma / \sqrt{2\phi_F - V_{BS}} \tag{5.126}$$

γ 是衬偏系数。

(3) 沟道电导 g_D 定义为

$$g_D \equiv \frac{\partial I_D}{\partial V_{DS}}\bigg|_{V_{GS},\,V_{BS}} \tag{5.127}$$

g_D 是 V_{GS}和 V_{BS}一定时 $I_D - V_{DS}$曲线的斜率,用电压控制模型有

$$g_D = \begin{cases} \dfrac{Z}{L}\mu_n C_{ox}(V_{GS} - V_T - V_{DS}) & \text{(非饱和区)} \\ 0 & \text{(饱和区)} \end{cases} \tag{5.128}$$

实际上,由于有效沟道长度随 V_{DS}改变(类似于双极晶体管中的 Early 效应)等原因,在饱和区 $g_D \neq 0$。

(4) 对 g_m 和 g_D 的修正 实际器件中,由于源区和漏区的体电阻以及欧姆接触电阻等原因,在沟道的源端和漏端存在电阻(这里分别以 R_S 和 R_D 表示),这些电阻使加在器件上的实际栅压 V'_{GS}和实际漏压 V'_{DS}减小,归根结底是器件的跨导和沟道电导受到影响。

我们可以通过图 5.11 所示的等效电路考虑 R_S 和 R_D 的影响,这里隐含着 R_S 和 R_D 与栅压无关的假定。如果施加在器件电极上的漏压和栅压分别为 V_{DS}和 V_{GS},并用 V'_{DS}和 V'_{GS}表示有效或本征漏压和栅压,则有

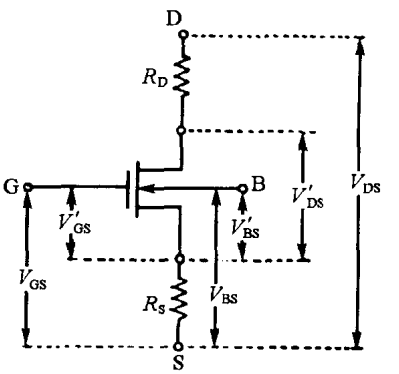


图 5.11 具有源电阻和漏电阻的 MOSFET

$$V'_{GS} = V_{GS} - I_D R_S \tag{5.129}$$

$$V'_{DS} = V_{DS} - I_D(R_S + R_D) \tag{5.130}$$

假定 V_{BS}不变,则漏极电流的微分方程为

$$\begin{aligned} dI_D &= \frac{\partial I_D}{\partial V'_{GS}}\bigg|_{V'_{DS}} dV'_{GS} + \frac{\partial I_D}{\partial V'_{DS}}\bigg|_{V'_{GS}} dV'_{DS} \\ &= g'_m\, dV'_{GS} + g'_D\, dV'_{DS} \end{aligned} \tag{5.131}$$

将分别由(5.129)和(5.130)得到的 $dV'_{GS}=dV_{GS}-R_S dI_D$ 和 $dV'_{DS}=dV_{DS}-(R_S+R_D)dI_D$ 代入上式,可以得到具有串联电阻的 MOSFET 的实际跨导和沟道电导分别为

$$g_m=\frac{g'_m}{1+R_S g'_m+(R_S+R_D)g'_D} \tag{5.132}$$

$$g_D=\frac{g'_D}{1+R_S g'_m+(R_S+R_D)g'_D} \tag{5.133}$$

这说明,由于串联电阻的存在,器件实际的 g_m 和 g_D 减小了。特别是在短沟道器件中,为了减小沟道电场以减小热载流子效应,通常采用轻掺杂漏(LDD)结构,在漏端引入高阻区,所以串联电阻的影响尤其突出。

3. 本征 MOSFET 的小信号等效电路

MOSFET 各端电压的微小变化引起漏极电流的变化可表示为

$$\begin{aligned} dI_D&=\frac{\partial I_D}{\partial V_{GS}}dV_{GS}+\frac{\partial I_D}{\partial V_{DS}}dV_{DS}+\frac{\partial I_D}{\partial V_{BS}}dV_{BS}\\ &=g_m dV_{GS}+g_D dV_{DS}+g_{mB}dV_{BS} \end{aligned} \tag{5.134}$$

在交流小信号的情形下,如果器件能跟上交流电压的变化,则式(5.134)中的微分量可用相应的交流量代替,所以漏极电流的交流分量为

$$\tilde{i}_D=g_m\tilde{V}_{GS}+g_D\tilde{V}_{DS}+g_{mB}\tilde{V}_{BS} \tag{5.135}$$

上式表示在小信号情形下,交流漏极电流是电流源 $g_m\tilde{V}_{GS}$, $g_{mB}\tilde{V}_{BS}$ 和 $g_D\tilde{V}_{DS}$ 共同作用的结果。

根据上述,本征 MOSFET 的交流小信号等效电路包括 5 个电容和 3 个电流源,如图 5.12 所示。

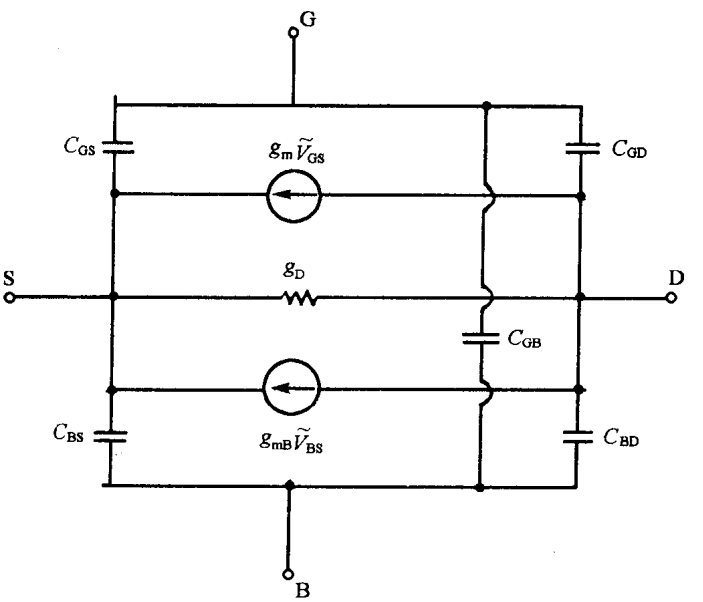


图 5.12 本征 MOSFET 的小信号等效电路

实际使用时，MOSFET 通常采用共源连接，以 G 和 S 作输入端，D 和 S 作输出端，并且 B 和 S 短路。这时，交流小信号等效电路如图 5.13。

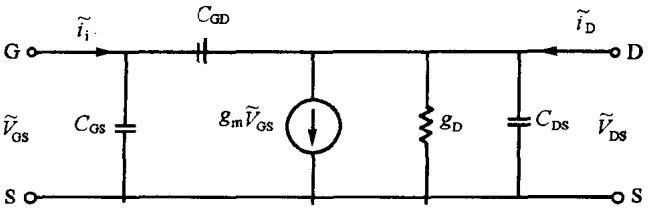


图 5.13　MOSFET 共源连接的小信号等效电路

4. 截止频率 f_T

定义 f_T 为输出端交流短路情形下 MOSFET 的输出电流 $\tilde{i}_D$ 和输入电流 $\tilde{i}_i$ 相等时的频率。根据图 5.13，输出短路时有

$$\tilde{i}_i = 2\pi f(C_{GS} + C_{GD})\,\tilde{V}_{GS} \simeq 2\pi f(ZLC_{ox})\,\tilde{V}_{GS} \tag{5.136}$$

$$\tilde{i}_D = g_m\,\tilde{V}_{GS} \tag{5.137}$$

令 $|\tilde{i}_D/\tilde{i}_i| = 1$，得到

$$f_T = \frac{g_m}{2\pi ZLC_{ox}} \tag{5.138}$$

用电压控制模型有

$$f_T = \begin{cases} \dfrac{\mu_n}{2\pi L^2}V_{DS} & \text{（非饱和区）} \\ \dfrac{\mu_n}{2\pi L^2}(V_{GS} - V_T) & \text{（饱和区）} \end{cases} \tag{5.139}$$

另一方面，载流子通过沟道的渡越时间 τ 是描述器件开关速度的一种量度。τ 越小，器件的开关速度或响应速度（响应带宽）越大。τ 和 g_m 及 C_G 之间有如下关系（见(4.69)式）：

$$\frac{g_m}{C_G} = \frac{1}{\tau}$$

作为一级近似，本征栅电容 C_G 就是间隔为 d_{ox}、面积为 ZL 的平行板电容，因此

$$\frac{g_m}{C_G} = \frac{\mu_n}{L^2}(V_{GS} - V_T) \tag{5.140}$$

其中利用了式(5.124)来计算饱和区的跨导 g_m。

由式(5.139)和式(5.140)可知，要提高 MOSFET 的截止频率或工作速度，必须缩短沟道长度 L 和提高沟道载流子的迁移率 μ_n。显然，与 p 沟 MOSFET 相比，n 沟器件有较高的截止频率和较快的工作速度。由于存在寄生效应（例如源/漏区的覆盖电容），实际器件的工作速度比从上述公式预期的要低得多。

5.2.6　短沟道效应

小型化一直是半导体器件发展的方向及先进性的标志。由于微细加工技术，特别是高分辨

率电子和 X 射线刻蚀技术的发展，场效应晶体管的沟道长度已减小到 0.5μm 以下，这种亚微米尺寸的线条宽度正是目前超大规模集成电路(VLSI)的基本特征。随着沟道长度的减小，沟道区的二维电势分布和强电场使器件性能偏离长沟道特性十分明显，这种偏离就是短沟道效应。短沟道效应使器件工作复杂化，并且使器件性能变坏，应该消除这些效应，或使之减至最小，从而使几何上的短沟道器件保持长沟道特性。

1. 沟道长度调制效应

如以前所述，当 $V_{DS} > V_{Dsat}$ 时，沟道在漏端附近“夹断”，漏极电流处于饱和状态。夹断点或速度饱和点(图 5.14 中的 P 点)仍然有饱和电压 V_{Dsat}，超过 V_{Dsat} 部分的漏电压施加在从 P 点到 n^+ 漏区的夹断区上，夹断区长度为 l_d。V_{DS} 增加时，l_d 必须增加以承受更高的过剩电压($V_{DS} - V_{Dsat}$)，这样，有效沟道长度必须减小。称这种效应为沟道长度调制。由于有效沟道长度随 V_{DS} 增加而减小，饱和区漏极电流 I_D 将随 V_{DS} 增加而并不“饱和”；而且，沟道越短，I_D 随 V_{DS} 增加的速率越大。这一现象是场效应晶体管研究中最早认识到的短沟道效应。

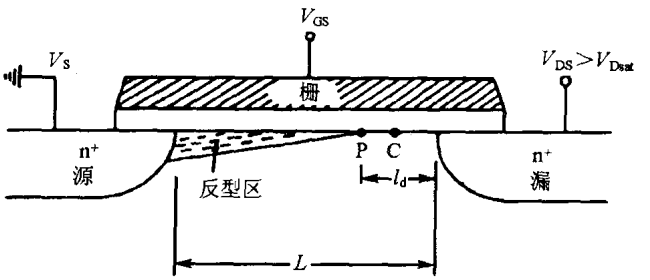


图 5.14　工作在饱和区的 MOSFET 示意图(其中 P 是夹断点)

在耗尽近似中，对 n 沟 MOSFET，l_d 区域内的电荷由体电荷密度为 qN_A 的电离受主组成。取坐标 y' 沿沟道方向，原点为 P 点。在 P 点($y'=0$)，沿 y' 方向的电场强度记作 $\mathscr{E}_p$，并假定 l_d 区域内紧靠表面处的电力线是水平的，则可以应用一维泊松方程

$$\begin{cases} \dfrac{d^2V}{dy'^2} = \dfrac{qN_A}{\varepsilon_s} \\ \left.\dfrac{dV}{dy'}\right|_{y'=0} = \mathscr{E}_p \end{cases} \tag{5.141}$$

解之可得

$$V_{DS} - V_{Dsat} = \frac{qN_A}{2\varepsilon_s} l_d^2 + \mathscr{E}_p l_d \tag{5.142}$$

所以

$$l_d = \sqrt{\frac{V_{DS} - V_{Dsat}}{\alpha} + \left(\frac{\mathscr{E}_p}{2\alpha}\right)^2} - \frac{\mathscr{E}_p}{2\alpha} \tag{5.143}$$

$$\alpha = \frac{qN_A}{2\varepsilon_s}$$

$\mathscr{E}_p$ 最终按经验确定，以得到同实验结果的最佳吻合。

$V_{DS} = V_{Dsat}$ 时；漏极电流的表示式在前面已经给出，为

$$I_{Dsat} = \frac{Z}{2L}\mu_n C_{ox}(V_{GS} - V_T)^2 \tag{5.144}$$

$V_{DS} > V_{Dsat}$时，有效沟道长度 $L - l_d$ 上的情形和 $V_{DS} = V_{Dsat}$时相同，只是应当用 $L - l_d$ 代替式(5.144)中的 L，故有

$$I_D = \frac{I_{Dsat}}{1 - l_d/L} \tag{5.145}$$

上式表明，随着 l_d(或 V_{DS})增加，漏极电流也增加。亦即在考虑沟道长度调制效应的情形下，饱和区漏极电流并不饱和。一般，当 l_d 远小于 L 时，作为一级近似，方程(5.145)为

$$I_D = I_{Dsat}\left(1 + \frac{l_d}{L}\right) \tag{5.146}$$

实际分析中，往往采用更简单的模型(参见图 5.15)，这就是假设 I_D 与 $V_{DS} - V_{Dsat}$的关系是线性的，即

$$I_D = I_{Dsat}\left[1 + \frac{V_{DS} - V_{Dsat}}{V_A}\right] \tag{5.147}$$

V_A 与双极型晶体管中的 Early 电压相当，由 MOSFET 的输出特性确定。V_A 越小，沟道长度调制越显著，漏端附近容易出现复杂的电场分布。为了减小这种效应，沟道区掺杂浓度 N_A 不能太低。

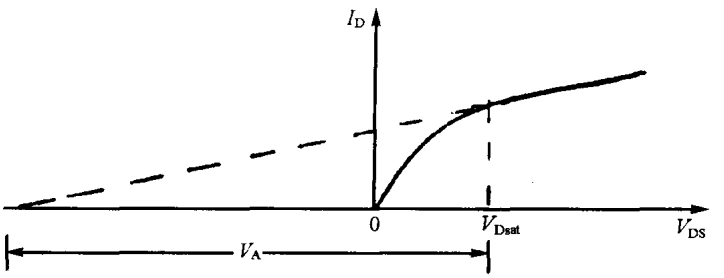


图 5.15　饱和区电流的线性模型

2. *阈值电压的短沟效应和窄沟效应*

在电压控制模型中，曾简单假定栅电场引起的耗尽区为矩形，耗尽层电荷为 $|Q_b| = qN_AX_d$。该假定忽略了源衬、漏衬 p-n 结耗尽区的影响。实际上，三者的耗尽区是重叠在一起的，因此有效栅控电荷 Q_b'将小于 Q_b。当沟长缩小时，栅控电荷减少了 $Q_L = Q_b - Q_b'$，从而导致阈值电压降低。为了简化，人们常用“电荷分享”模型来得到有效阈值电压 V_T'的解析(或半经验的)结果。电荷分享模型把沟道耗尽区分为两部分，一部分由栅压引起，另一部分由源衬、漏衬 p-n 结引起。这样便可以得到 Q_b'，只要用 Q_b'代替阈值电压表示式中的 Q_b，即得到有效阈值电压 V_T'。在栅下耗尽区均匀的假定下，有

$$V_T' = V_{FB} + 2\phi_F - \frac{Q_b'}{C_{ox}} = V_{FB} + 2\phi_F + \frac{Q_b'}{Q_b}\gamma\sqrt{2\phi_F - V_{BS}} \tag{5.148}$$

其中 γ 是衬偏系数。

图 5.16 表示一个极其简单的电荷分享模型。首先它假定源漏之间的偏压为零，即 $V_{DS} = 0$；其次假定源衬、漏衬 p-n 结的自建电压 V_{bi}等于 $2\phi_F$，故三个耗尽区的宽度相等，皆为

$$X_d = \sqrt{\frac{2\varepsilon_s}{qN_A}(2\phi_F - V_{BS})} \tag{5.149}$$

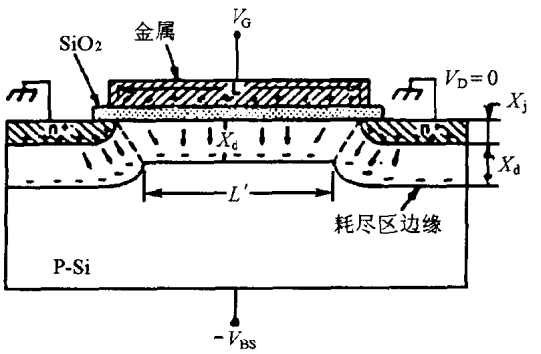


图 5.16　推导栅控电荷 Q_b'的梯形近似

Q_b'由图中的梯形面积决定。假定 n^+ 区的边界为圆柱形，其半径等于结深 X_j，则经过简单的几何分析，就可以得到

$$\frac{Q_b'}{Q_b} = 1 - \frac{X_j}{L}\left(\sqrt{1 + \frac{2X_d}{X_j}} - 1\right) \tag{5.150}$$

比值 Q_b'/Q_b（称为电荷分享因子）描述了沟道区中栅控耗尽电荷在总耗层电荷中所占的比例，它总小于 1。显然，长沟器件的电荷分享因子接近 1，即 Q_b'与 Q_b 接近。

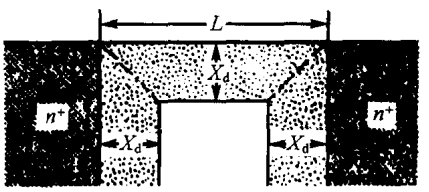


图 5.17　结深 X_j 很大时栅控电荷的梯形近似

若 X_d/X_j 较小，则可以把式(5.150)括号中的量展开成级数，并略去高次项，结果为

$$\frac{Q_b'}{Q_b} \simeq 1 - \frac{X_d}{L} \tag{5.151}$$

图 5.17 表示这种极限情形。X_d/X_j 较大时，近似式(5.151)有相当大的误差，因此引入一个经验参数 α_1，把它的适用范围扩大($\alpha_1 > 1$)：

$$\frac{Q_b'}{Q_b} = 1 - \alpha_1 \frac{X_d}{L} \tag{5.152}$$

将式(5.152)代入式(5.148)，有

$$V_T' = V_{FB} + 2\phi_F + \left(1 - \alpha_1 \frac{X_d}{L}\right)\gamma \sqrt{2\phi_F - V_{BS}} \tag{5.153}$$

短沟道效应引起阈值电压的改变量为

$$\Delta V_T = V_T - V_T' = \alpha_1 \frac{X_d}{L}\gamma \sqrt{2\phi_F - V_{BS}} = 2\alpha_1 \frac{\varepsilon_s}{\varepsilon_{ox}} \frac{d_{ox}}{L}(2\phi_F - V_{BS}) \tag{5.154}$$

为了降低阈值电压的短沟道效应，栅氧化层厚度 d_{ox} 应尽可能小，以加强栅压对沟道区耗尽电荷的控制。

当 $V_{DS} \neq 0$ 时，我们采用如图 5.18 所示的模型来近似分析，有

$$\frac{Q_b'}{Q_b} \simeq 1 - \alpha_1 \frac{1}{L} \frac{y_S + y_D}{2} \tag{5.155}$$

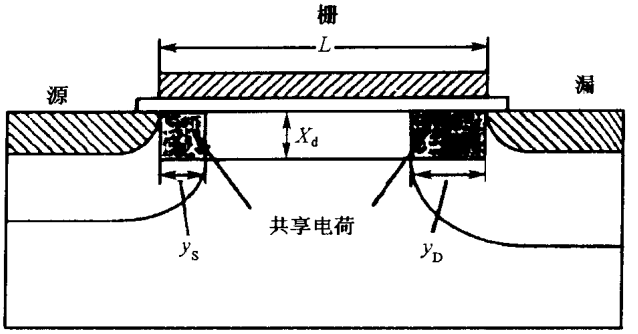


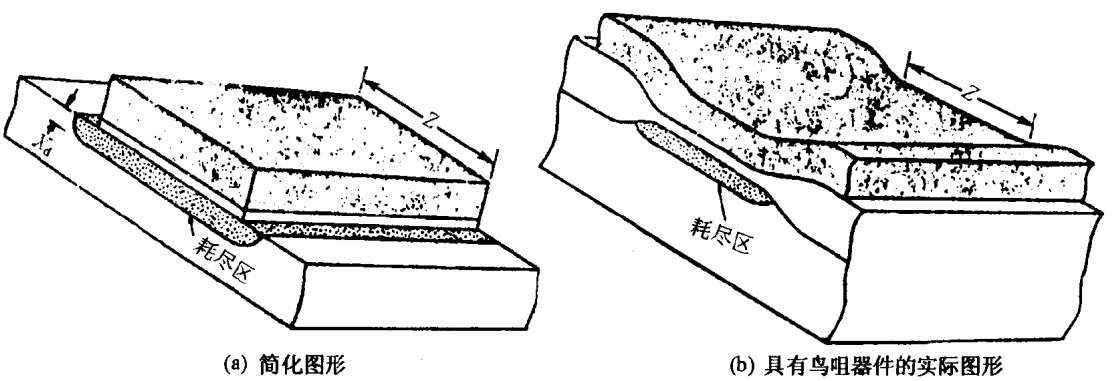
图 5.18　$V_{DS}\neq 0$ 时，栅下耗尽区电荷与源、漏及栅的关系

$\alpha_1 \geqslant 1$ 是经验参数；y_S 和 y_D 分别为源衬、漏衬 p-n 结的水平耗尽区宽度，它们分别由公式(5.92)和(5.93)表示。受 V_{DS}影响，阈值电压的减小量为

$$\Delta V_T = \frac{\alpha_1(y_S + y_D)\sqrt{\varepsilon_s q N_A \phi_F}}{LC_{ox}} \tag{5.156}$$

为简单起见，写出上式时假定了源和衬底短路(即 $V_{BS}=0$)。当 V_{DS}不很大时，ΔV_T 随 V_{DS}增加，即有效阈值电压 V_T'随 V_{DS}的增加而减小。若 V_{DS}很大，以致 $y_S + y_D = L$，则源结和漏结的耗尽区连通，器件处于穿通状态，有很大的漏极电流，并且不受栅压控制。

窄沟效应　沟道很窄时，也会影响器件的阈值电压。一个器件沿其沟道宽度方向(Z 方向)的截面示于图 5.19(a)。为简单起见，此图画得非常理想化。一个实际器件的沟道截面因某些制造工艺看上去可能如图 5.19(b)，在厚场氧化层(场氧)与薄栅氧化层(有源区)的过渡区形成了类似鸟咀的结构。从两幅图中都可看出，耗尽区并不正好限于薄栅氧化层下面的区域内。这是因为从栅电荷出发的电力线有一些终止在沟道两侧的电离受主上。这些电力线构成了所谓边缘场。如果 Z 较大，则耗尽区的两侧部分占总耗尽区体积的百分数较小，因而可以忽略；然而，当 Z 只有几个微米时，两侧部分占总体积的百分数就变大了。与短沟道情况相反，现在栅要负责耗尽一个大于长沟理论所预计的矩形区域。于是，为了在反型层形成之前，耗尽那样体积的区域，V_{GS} 必须更大，即器件应该有较大的有效阈值电压 V_T'。



(a) 简化图形　　(b) 具有鸟咀器件的实际图形

图 5.19　边缘场引起耗尽区侧向扩展

如图 5.19(a)，假定侧向扩展为圆柱形，并且 $V_{DS}=0$，则不难得到

$$\frac{Q_b'}{Q_b} = 1 + \frac{\pi}{2}\frac{X_d}{Z} \tag{5.157}$$

X_d 为栅下耗尽层厚度，Q_b'为栅控电荷($Q_b'>Q_b$)。窄沟时阈值电压的增加量为

$$\Delta V_T = \frac{\pi}{2}\frac{X_d}{Z}\gamma\sqrt{2\phi_F - V_{BS}} = \pi\frac{\varepsilon_s}{\varepsilon_{ox}}\frac{d_{ox}}{Z}(2\phi_F - V_{BS}) \tag{5.158}$$

为了减小阈值电压的窄沟道效应，同样应减小栅氧化层厚度 d_{ox}，以避免栅电荷的电力线超出栅极近缘侧向扩展。

3. 漏感应势垒降低(DIBL)效应

当 L 减小、V_{DS}增加时，漏源耗尽区越来越靠近，引起电力线从漏到源的穿越，使源端势垒降低，从源区注入沟道的电子增加，导致漏源电流增加。通常称该过程为漏感应势垒降低，简写为 DIBL。图 5.20 是 $V_{DS}=0$ 和 $V_{DS}>0$ 时 n 沟 MOSFET 沟道表面的能带和电势分布的示意图。图中，$V_s(0,y)$是相对 n^+ 源区($y=0$)的表面势，y_S 和 y_D 分别为源结和漏结的水平耗尽区宽度。$V_{DS}=0$ 时，在 $y=y_S$ 至 $y=L-y_S-y_D$ 的区域(实际的沟道区)内，$V_S=V_S^0$(常数)，y_S 处的势垒高度为 $-qV_S^0$。$V_{DS}>0$ 时，沟道表面势增加 $V(y)$，结果电势最小处 y_{min}(y_S 附近)的电势为 $V_S(y_{min})=V_S^0+V(y_{min})$，增加 $V(y_{min})$，源和沟道之间的势垒则相应降低了 $qV(y_{min})$，这就是漏感应势垒降低。显然，对一定的 V_{DS}，器件的 L 越小，DIBL 越显著，漏极电流的增加越显著，以致器件不能关断。所以，DIBL 效应是对 MOS 器件尺寸缩小的一个基本限制。

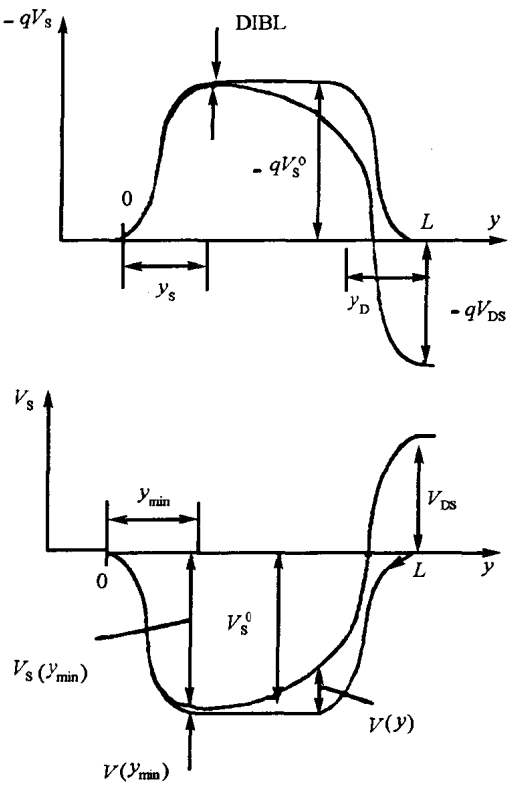


图 5.20　n 沟 MOSFET 沟道表面的能带图和电势分布
(对称分布对应 $V_{DS}=0$，非对称分布对应 $V_{DS}>0$)

如上所述，DIBL 是漏电压 V_{DS}引起的沿沟道方向的电势分布使源和沟道间的势垒降低 $qV(y_{min})\simeq qV(y_S)$。所以，为了分析 DIBL，必须既考虑受栅偏压和衬底偏压控制的垂直电场 $\mathscr{E}_x$，也考虑受漏偏压控制的沿沟道方向的电场 $\mathscr{E}_y$，即必须求解二维泊松方程。为了简化分析，我们假定器件的工作状态是在亚阈值区。

首先，对于亚阈值区，沟道区表面电子浓度很低，可以采用耗尽近似，二维泊松方程为

$$\frac{\partial \mathscr{E}_x}{\partial x}+\frac{\partial \mathscr{E}_y}{\partial y}=-\frac{qN_A}{\varepsilon_s} \tag{5.159}$$

固定 y，将方程(5.159)在耗尽区上对 x 积分，得到

$$-\mathscr{E}_s+\frac{\partial \mathscr{E}_y}{\partial y}X_d=-\frac{qN_A}{\varepsilon_s}X_d \tag{5.160}$$

X_d 是耗尽区厚度，$\partial\mathscr{E}_y/\partial y$ 应理解为对 X_d 的平均值。X_d 可表示为

$$X_d=\sqrt{\frac{2\varepsilon_s}{qN_A}(V_s+V_{bi})} \tag{5.161}$$

$\mathscr{E}_s$ 是沟道区表面电场的垂直分量，由电势平衡方程有

$$\mathscr{E}_s=\frac{C_{ox}}{\varepsilon_s}(V_{GS}-V_{FB}-V_s-V_{bi}) \tag{5.162}$$

V_s 是相对 n^+ 源区(源接触)的表面势，为

$$V_s(y)=V_s^0+V(y) \tag{5.163}$$

除源、漏端附近外，可假设 $V_{DS}=0$ 时沟道区内 $\partial\mathscr{E}_y/\partial y=0$。这样，$V_{DS}=0$ 时方程(5.160)应为

$$-\mathscr{E}_s^0=-\frac{qN_A}{\varepsilon_s}X_d^0 \tag{5.164}$$

其中 X_d^0 和 $\mathscr{E}_s^0$ 分别为 $V=0$(即 $V_{DS}=0$)时的耗尽区厚度和表面电场的垂直分量。

从式(5.160)减去式(5.164)，得到

$$\frac{\partial^2 V(y)}{\partial y^2}X_d-\frac{C_{ox}}{\varepsilon_s}V(y)=\frac{qN_A}{\varepsilon_s}X_d^0\left[\sqrt{1+\frac{V(y)}{V_s^0+V_{bi}}}-1\right] \tag{5.165}$$

除源、漏端附近外，$V(y)$很小，上式中的平方根项可以展开成 $V(y)$的幂级数，并保留至线性项，得到

$$\frac{\partial^2 V}{\partial y^2}-\frac{V}{\lambda^2}=0 \tag{5.166}$$

其中，

$$\lambda=X_d^0\left(1+\frac{\varepsilon_{ox}}{\varepsilon_s}\frac{X_d^0}{d_{ox}}\right)^{-1/2} \tag{5.167}$$

假定方程(5.166)的解也适合于源结的水平耗尽区，即 $y<y_S$ 的区域，这并不会带来多大误差。这样，就有边界条件 $V(y=0)=0$；与之相应，方程(5.166)的解为

$$V(y)=V_0\sinh\left(\frac{y}{\lambda}\right) \tag{5.168}$$

在式(5.168)中，V_0 是待定常数，可以由附加表面势引起耗尽区电荷增量 $\Delta Q_B'$(栅电荷相应增加 $-\Delta Q_B'$)确定：

$$\Delta Q_B' = -\int_{y_S}^{L-y_D} C_D V(y)\mathrm{d}y \simeq \int_{y_S}^{L-y_D} \frac{\varepsilon_s}{X_d^0} V_0 \sinh\left(\frac{y}{\lambda}\right)\mathrm{d}y \tag{5.169}$$

C_D是沟道耗尽层电容。上式用到了缓变沟道近似。另一方面，根据电荷共享原理，$\Delta Q_B'$只是V_{DS}在栅下引起电荷总量ΔQ的一部分，即$k\Delta Q$，比例系数k为0.5左右。ΔQ可以通过图5.21粗略估计为(图中栅下衬底内阴影区面积)：

$$\Delta Q \simeq qN_A Z(y_D^2 - y_D^{0\,2}) \tag{5.170}$$

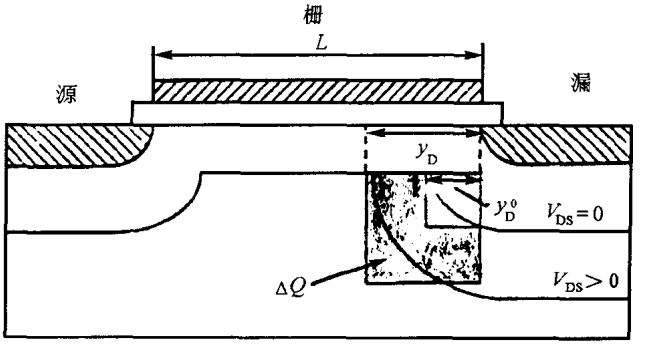


图5.21　漏偏压V_{DS}在栅下衬底内感生电荷总量ΔQ的简化模型

$y_D^0 = y_S$是$V_{DS}=0$时漏结的水平耗尽区宽度。利用表示y_S和y_D的公式(5.92)和(5.93)，可得

$$\Delta Q_B' = k\Delta Q = 2kZ\varepsilon_s V_{DS} \tag{5.171}$$

完成式(5.169)的积分，并与式(5.171)比较，即得

$$V_0 = 2kV_{DS}\frac{X_d^0}{\lambda}\left[\cosh\left(\frac{L-y_D}{\lambda}\right) - \cosh\left(\frac{y_S}{\lambda}\right)\right]^{-1} \tag{5.172}$$

将式(5.172)代入式(5.168)，并令$y = y_{min} = y_S$，就得到漏感应势垒降低(DIBL)为

$$qV(y_S) = 2kq\frac{X_d^0}{\lambda}\frac{\sinh(y_S/\lambda)}{\cosh\left(\frac{L-y_D}{\lambda}\right) - \cosh\left(\frac{y_S}{\lambda}\right)}V_{DS} \tag{5.173}$$

上式表明，栅长L一定时，DIBL随漏压V_{DS}增加。当L很小或V_{DS}很大以致$y_S + y_D = L$时，上式不再成立，这对应器件的穿通状态。

假定理想因子η不随偏压而改变，则由方程(5.90)可以得到表面势的变化为

$$\Delta V_s = -\frac{\Delta V_T}{\eta} \tag{5.174}$$

作为势垒降低的结果，漏偏压V_{DS}将使阈值电压下降。由式(5.174)及(5.173)，有

$$\Delta V_T = -\eta\Delta V_s(y_S) = -\eta V(y_S) = -\sigma V_{DS} \tag{5.175}$$

其中

$$\sigma = \frac{2k\eta X_d^0}{\lambda}\frac{\sinh(y_S/\lambda)}{\cosh\left(\frac{L-y_D}{\lambda}\right) - \cosh\left(\frac{y_S}{\lambda}\right)} \tag{5.176}$$

σ称为DIBL因子。当短沟器件工作在阈值电压附近时，DIBL效应非常严重。

4. 穿通效应

当漏结的耗尽区和源结的耗尽区连通(即 $y_S + y_D = L$)时,源和沟道间的势垒显著降低,注入沟道的电子浓度 n 很大,它们将以漂移方式通过源和漏之间的空间电荷区,电流密度可表示为

$$J_{sp} = -qn\mu_n\mathscr{E} \tag{5.177}$$

J_{sp}习惯上称为空间电荷限制电流。

设衬底掺杂浓度显著低于注入的电子浓度,则空间电荷区的泊松方程为

$$\frac{d\mathscr{E}}{dy} = -\frac{qn}{\epsilon_s} \tag{5.178}$$

所以

$$J_{sp} = \epsilon_s\mu_n\mathscr{E}\frac{d\mathscr{E}}{dy} \tag{5.179}$$

积分上式得到

$$\mathscr{E} = -\left[\frac{2J_{sp}}{\epsilon_s\mu_n}y + \mathscr{E}_0^2\right]^{1/2} \tag{5.180}$$

负号表示电场的方向由漏指向源;$\mathscr{E}_0$ 是沟道源端的电场强度,一般可以略去。

根据 $\mathscr{E} = -dV/dy$,对方程(5.180)积分,得到

$$J_{sp} = \frac{q\epsilon_s\mu_n V_{DS}^2}{8L^3} \tag{5.181}$$

J_{sp}并联于受栅压控制的沟道电流。所以,在穿通情形下,即使栅压低于阈值,源漏之间也会有电流流过。

5. 速度饱和与两区模型

至今为止,一直假定沟道电子的迁移率与沿沟道方向的电场 $\mathscr{E}_y$ 无关,即假定沟道内 $\mathscr{E}_y$ 总是比速度饱和的临界场强 $\mathscr{E}_c$ 低得多。实际上,对于一定的漏偏压 V_{DS}(例如 5V 工作电压标准),随着器件沟道长度 L 的减小,沿沟道的场强 $\mathscr{E}_y(y)$增强,电子迁移率将和 $\mathscr{E}_y$ 有关,其漂移速度甚至饱和。在这种情形下,基于恒定迁移率的电流理论必须修正。

(1) 速度饱和效应

对于电子漂移速度和电场的关系,我们采用下述经验公式:

$$v = \begin{cases} \dfrac{\mu_n\mathscr{E}_y}{1 + \mathscr{E}_y/\mathscr{E}_c} & \mathscr{E}_y < \mathscr{E}_c \\ v_s & \mathscr{E}_y > \mathscr{E}_c \end{cases} \tag{5.182}$$

其中,μ_n 是沟道电子的低场迁移率;$\mathscr{E}_c$ 是速度饱和临界场强,在 Si 中约为 5×10^4V/cm;v_s 为电子的饱和漂移速度,为理论饱和速度的一半($v_s = \mu_n\mathscr{E}_c/2$)。将 $\mathscr{E}_y = dV/dy$ 代入上式,得

$$v = \mu_n\frac{dV/dy}{1 + \dfrac{1}{\mathscr{E}_c}\dfrac{dV}{dy}} \tag{5.183}$$

由于迁移率不再与电场无关,即 v 不再与 $\mathscr{E}_y$ 成正比,漏极电流应表示成

$$I_D = Z(-Q_i)v \tag{5.184}$$

将 v 的表示式(5.183)代入,得

$$I_D\left[1 + \frac{1}{\mathscr{E}_c}\frac{dV}{dy}\right] = Z\mu_n(-Q_i)\frac{dV}{dy} \tag{5.185}$$

从 $y=0$ 至 $y=L$ 积分上式，并假定 μ_n 是常数，可得

$$I_D = \frac{Z}{L}\mu_n \int_0^{V_{DS}} (-Q_i)\mathrm{d}V \Big/ \left(1 + \frac{V_{DS}}{L\mathscr{E}_c}\right) \tag{5.186}$$

这一方程说明，与不考虑速度饱和效应相比，漏极电流应乘以因子$(1 + V_{DS}/L\mathscr{E}_c)^{-1}$。这样，对电压控制模型就有

$$I_D = \frac{Z}{2L}\mu_n C_{ox}(2V'_{GS} - V'^2_{DS}) \frac{1}{1 + \dfrac{V_{DS}}{L\mathscr{E}_c}} \tag{5.187}$$

其中 $V'_{GB} = V_{GS} - V_T$ 是夹断饱和电压。

当漏端($y=L$)的场强 $\mathscr{E}_y(L) = \mathscr{E}_c$ 时，电流开始饱和，其值为

$$I'_{Dsat} = \frac{Z}{2L}\mu_n C_{ox}(2V'_{GS} - V'_{Dsat}) \frac{1}{1 + V'_{Dsat}/L\mathscr{E}_c} \tag{5.188}$$

V'_{Dsat}表示速度饱和电压，右上角加撇以示同夹断饱和相区别。另一方面，I'_{Dsat}也可以通过饱和漂移速度 v_s 表示成

$$I'_{Dsat} = Zv_s C_{ox}(V'_{GS} - V'_{Dsat}) \tag{5.189}$$

由式(5.188)和(5.189)，并注意到 $v_s = \mu_n\mathscr{E}_c/2$，有

$$V'_{Dsat} = \frac{V'_{GS}L\mathscr{E}_c}{L\mathscr{E}_c + V'_{GS}} \tag{5.190}$$

结合(5.189)，(5.190)二式，有

$$I'_{Dsat} = Zv_s C_{ox} \frac{V'^2_{GS}}{L\mathscr{E}_c + V'_{GS}} \tag{5.191}$$

下面简短讨论速度饱和效应对器件 I_D-V_{DS}特性的影响。

① 若 L 很大以致 $L\mathscr{E}_c \gg V'_{GS}$(长沟道器件)，则有

$$V'_{Dsat} = V'_{GS},\quad I'_{Dsat} = (Z/2L)\mu_n C_{ox} V'^2_{GS}$$

这正是夹断饱和的情形。一般，$V'_{Dsat} < V_{Dsat}$，$I'_{Dsat} < I_{Dsat}$，也就是说，由于存在速度饱和效应，V_{DS}之值尚未达到夹断饱和电压 V'_{GS}时，器件的漏极电流已经饱和，其值比夹断饱和电流值要小，并且同栅压 V'_{GS}不再有平方律关系。

② 若 L 很小以致 $L\mathscr{E}_c \ll V'_{GS}$(短沟道器件)，则有

$$V'_{Dsat} = L\mathscr{E}_c,\quad I'_{Dsat} = Zv_s C_{ox} V'_{GS}$$

其中 $C_{ox}V'_{GS}$是沟道源端的电荷浓度。所以，对于 L 很小以致速度严重饱和的器件，饱和电流等于沟道电荷同饱和漂移速度相乘的乘积，而电荷数沿沟道的变化可以忽略。

(2) 两区模型(准二维分析)

随着 $\mathscr{E}_y(y)$增加，速度饱和临界电场 $\mathscr{E}_c$ 将在沟道内某一位置(图 5.22 中的 P 点)$y = L - l_d$ 达到。这样，器件的沟道区可分为两个区域，源端附近的区 I 内 $\mathscr{E}_c(y) < \mathscr{E}_c$，为非饱和区；漏端附近的区Ⅱ内 $\mathscr{E}_y(y) > \mathscr{E}_c$，为速度饱和区。如图 5.22 所示。

速度非饱和区$(0 < y < L - l_d)$

根据短沟道 MOSFET 的二维模拟结果，源端附近 $\mathscr{E}_y(y)$随位置 y 近似线性变化且比较缓慢，故在这一区域内可假设

$$\mathscr{E}_y(y) = \frac{\mathscr{E}_c}{L - l_d} y \tag{5.192}$$

对电流特性的分析，一般仍可采用缓变沟道近似，就像上面讨论速度饱和效应时所做的那样。也就是说，器件的漏极电流仍可表示成方程(5.188)或(5.191)的形式，只是式中的 L 用 $L - l_d$ 代替，而 l_d 则在对速度饱和区的分析以后确定。由于速度饱和点(P 点)可以随 V_{DS}的变化从漏端移入沟道内，其位置(即 l_d 的长度)取决于 $\mathscr{E}_y$ 等于临界值 $\mathscr{E}_c$ 的位置，因此上述两区模型适用于电流特性从线性区到饱和区的所有工作状态。同时，由于对速度饱和区将采用二维近似分析，这一模型也是基本适合于包括短沟道情形在内的关于 MOSFET 电流特性的近似模型。

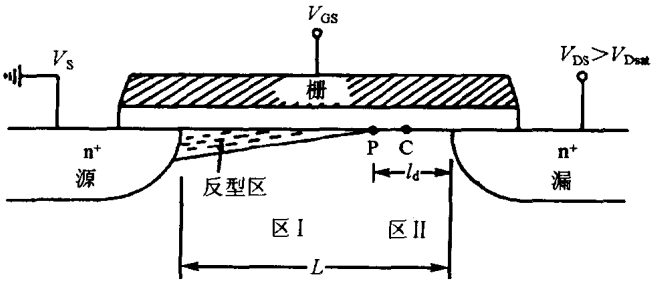


图 5.22　在饱和区计算短沟道器件电流的分区模型

速度饱和区($L - l_d < y < L$，或 $0 < y' < l_d$)

这一区域内 $\mathscr{E}_y(y)$很强，不能使用缓变沟道近似。准二维分析模型在这一区域主要体现在一定近似条件下对图 5.23 所示的 ABCD 区域应用高斯定理(高斯面的外法线方向与坐标方向相同时取“+”号，相反时取“−”号)：

$$-\mathscr{E}_c X_j + \mathscr{E}_y(y') X_j - \int_0^{y'} \frac{\varepsilon_{ox}}{\varepsilon_s} \mathscr{E}_{ox}(y')\mathrm{d}y' = -\frac{q(N_A + N_I)}{\varepsilon_s} X_j\, y' \tag{5.193}$$

其中 $y' = y - (L - l_d)$；N_I 是饱和点处的反型载流子浓度；N_A 是沟道内电离受主浓度；$\mathscr{E}_y(y')$是沿沟道方向的场强，假定它与深度(结深)X_j 无关；$\mathscr{E}_{ox}(y')$是垂直于 Si/SiO_2 界面的电场，表示式为

$$\mathscr{E}_{ox}(y') = \frac{V_{GS} - V_{FB} - 2\phi_F - V(y')}{d_{ox}} \tag{5.194}$$

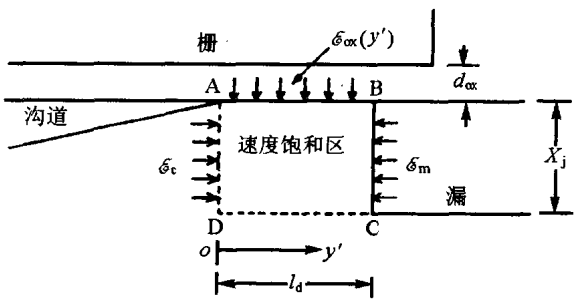


图 5.23　分析速度饱和区特性的示意图

其中 $\psi_s = 2\phi_F$ 是强反型时的表面势，即半导体表面层上的电压；$V_{GS} - V_{FB} - V(y')$是 y'处超过平带电压的有效栅压，它等于表面势和表面层电荷在栅氧化层上产生的电压之和；d_{ox}是栅氧化层厚度。需要注意，准二维分析模型是为了得到描述十分复杂的二维现象的简单的半经

验公式而提出的，不可能推导过程中的每一步都是严格的。实际上，在写出方程(5.193)时假定了通过 DC 线的电场可以忽略，还假定了有效界面电荷 Q_0 等于 0。

方程(5.193)对 y' 微商，得到

$$\frac{d\mathscr{E}(y')}{dy'}X_j - \frac{\varepsilon_{ox}}{\varepsilon_s}\mathscr{E}_{ox}(y') = -\frac{q(N_A + N_I)}{\varepsilon_s}X_j \tag{5.195}$$

在 $y'=0$ 处，$d\mathscr{E}(y')/dy'$ 之值可以根据电场梯度连续的条件从式(5.192)得到，但为了简化结果，这里假定它为 0。于是对于 $y'=0$，可根据方程(5.196)写出

$$-\frac{\varepsilon_{ox}}{\varepsilon_s}\mathscr{E}_{ox}(0) = -\frac{q(N_A + N_I)}{\varepsilon_s}X_j \tag{5.196}$$

其中

$$\mathscr{E}_{ox}(0) = \frac{V_{GS} - V_{FB} - 2\phi_F - V'_{Dsat}}{d_{ox}} \tag{5.197}$$

方程(5.195)和(5.196)相减，得到

$$\frac{d\mathscr{E}_y(y')}{dy'} + \frac{V(y') - V'_{Dsat}}{l^2} = 0$$

或

$$\frac{d^2V(y')}{dy'^2} - \frac{V(y') - V'_{Dsat}}{l^2} = 0 \tag{5.198}$$

$$l = \sqrt{\frac{\varepsilon_s}{\varepsilon_{ox}}d_{ox}X_j} \tag{5.199}$$

边界条件为

$$\begin{aligned} V(y' = 0) &= V'_{Dsat} \\ \left.\frac{dV(y')}{dy'}\right|_{y'=0} &= \mathscr{E}_c \end{aligned} \tag{5.200}$$

方程(5.198)满足边界条件的解为

$$V(y') = V'_{Dsat} + l\mathscr{E}_c\sinh(y'/l) \tag{5.201}$$

电场强度 $\mathscr{E}_y(y')$ 和位置 y' 的关系为

$$\begin{aligned} \mathscr{E}_y(y') &= dV(y')/dy' \\ &= \mathscr{E}_c\cosh(y'/l) \end{aligned} \tag{5.202}$$

漏端($y' = l_d$)的场强最大。

利用双曲函数的下述性质：

$$\sinh x = \frac{e^x - e^{-x}}{2}, \quad \cosh x = \frac{e^x + e^{-x}}{2}$$

$$\cosh^2 x - \sinh^2 x = 1$$

在方程(5.201)和(5.202)中，令 $y' = l_d$ 和 $V(l_d) = V_{DS}$，可以得到漏端的场强 $\mathscr{E}_m$ 和速度饱和区长度 l_d，分别为

$$\mathscr{E}_m = \left[\left(\frac{V_{DS} - V'_{Dsat}}{l}\right)^2 + \mathscr{E}_c^2\right]^{1/2} \tag{5.203}$$

$$l_d = l\ln\left[\frac{V_{DS} - V'_{Dsat}}{l\mathscr{E}_c} + \frac{\mathscr{E}_m}{\mathscr{E}_c}\right] \tag{5.204}$$

与夹断饱和类似，$V_{DS} > V'_{Dsat}$时区域 $L - l_d$ 上的情形和 $V_{DS} = V'_{Dsat}$时区域 L 上的情形相同，故在速度饱和临界电流 I'_{Dsat}的表示式中将 L 用 $L - l_d$ 替换，就可以得到漏极电流 I_D 的表示式，显然它是 l_d 的函数。用 I'_{Dsat}表示，则有

$$I_D(l_d) = I'_{Dsat} \frac{L\mathscr{E}_c + V'_{Dsat}}{(L - l_d)\mathscr{E}_c + V'_{Dsat}} \tag{5.205}$$

器件的输出电阻($R_o = dV_{DS}/dI_D$)为

$$R_o = \frac{1}{I_D}\mathscr{E}_m \frac{(L - l_d)\mathscr{E}_c + V'_{Dsat}}{\mathscr{E}_c} \simeq \frac{1}{I_D}\frac{V_{DS} - V'_{Dsat}}{l}\left(L + \frac{V'_{Dsat}}{\mathscr{E}_c}\right) \tag{5.206}$$

根据上式，短沟道 MOSFET 饱和区的输出电阻随 V_{DS}的增加而增加。但在实际器件中，特别是在 n MOST 中，当 V_{DS}达到一定值时，R_o 不再增加，此后进一步增加 V_{DS}，R_o 将降低。R_o 降低是由于热电子效应和 DIBL 效应，它们使漏极电流 I_D 增大。

我们特别感兴趣的是 $\mathscr{E}_m \gg \mathscr{E}_c$ 的情形。对于一定的工作电压标准，V_{DS}一定，因此栅长 L 越小时沿沟道方向的电场越强，越是容易发生这种情形。这时，由于漏端附近电场很强，沟道载流子将获得十分大的动能，成为热电子。热电子可以引起碰撞电离，导致很大的衬底电流；还可以越过 Si/SiO_2 的界面，损伤栅氧化层，从而导致器件性能变坏。当 $\mathscr{E}_m \gg \mathscr{E}_c$ 时，式(5.203)简化为

$$\mathscr{E}_m \simeq \frac{V_{DS} - V'_{Dsat}}{l} \tag{5.207}$$

上式表明，如果假定速度饱和区的电场强度是常数且等于 $\mathscr{E}_m$，则 l 就是饱和区的长度。所以，l 是反映速度饱和区长度的特征量，它和栅氧化层厚度 d_{ox}、源区及漏区的扩散结深 X_j 有关，这已由公式(5.199)给出。但通常采用下述从实验得出的经验公式：

$$l \simeq \begin{cases} 0.22 d_{ox}^{1/3} X_j^{1/3} & (d_{ox} > 150\text{Å}) \\ 0.017 d_{ox}^{1/8} X_j^{1/3} L^{1/5} & (d_{ox} < 150\text{Å}, L < 0.5\mu\text{m}) \end{cases} \tag{5.208}$$

理论结果和实验结果存在明显差别，主要是由于理论分析采用的模型过于简单。另方面，经验公式因研究者而异，依赖于实验条件和工作经验。

6. 热电子效应

对于沟道长度很短的器件，一定的漏源电压 V_{DS}可以使漏端附近形成相当高的电场。随着载流子(对 n 沟 MOSFET 为电子)从源向漏移动，它们在漏端高电场区将得到十分大的动能，并且具有显著高于晶体中电子正常热能的能量，成为热电子。热电子可以通过碰撞电离产生次级电子-空穴对，其中电子(包括原始和次级电子)流入漏极，形成漏极电流 I_D；碰撞产生的次级空穴将流入衬底，形成衬底电流 I_b(见图 5.24)。通过测量 I_b 可以很好地监控沟道热电子和漏区电场的情况。当较大的衬底电流通过衬底时，会在衬底上产生电压降；由于 MOSFET 的源通常接地，该电压降将使源-衬结正偏，从而形成一个源-衬底-漏(n-p-n)结构的寄生双极晶体管，与 MOSFET 并联。这种复合结构(MOSFET 和寄生双极晶体管)是大多数短沟 MOSFET 导致漏源击穿的原因，并且还会引起 I-V 曲线的回滞现象；在 CMOS 电路中，则会导致闩锁效应。有一些具有足够高能量的热电子还可以越过 Si/SiO_2 界面势垒(对于电子约为 3.2eV，对于空穴约为4.9eV)，注入到栅氧化层中，它们中的大多数将被栅极收集，形成栅电流 I_G。热电子注入到栅氧化层中会引起界面陷阱，经过一段时间的积累，器件性能将会退

化，主要表现在阈值电压漂移、跨导降低和亚阈值斜率增加。这些退化将严重影响超大规模MOS集成电路的可靠性。所以，热电子使器件性能退化主要和栅电流 I_G 有关，而不是与衬底电流 I_b 有关，尽管 I_G 比 I_b 要小几个数量级。

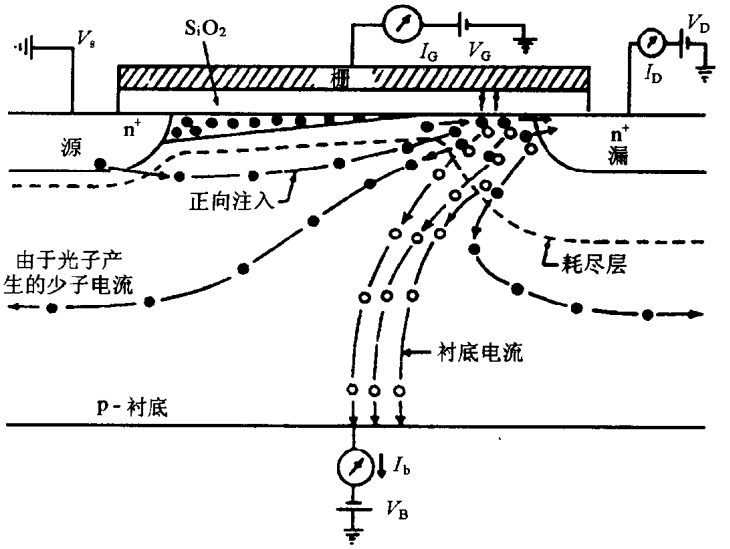


图 5.24　描述 nMOST 热电子效应的示意图

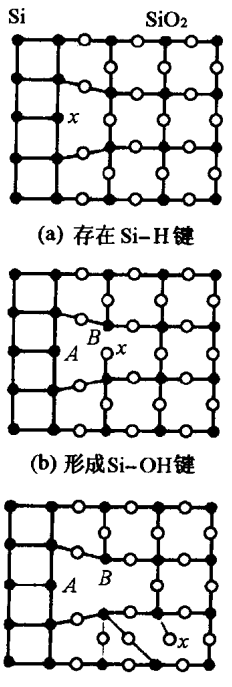


图 5.25　SiO_2 界面的二维假想模型
●: Si　○: O　x: H

I_G 及 I_b 同 I_D 的关系可表示如下：

$$I_b = C_1 I_D e^{-E_i/q\mathscr{E}\lambda} \tag{5.209}$$

$$I_G = C_2 I_D e^{-E_b/q\mathscr{E}\lambda} \tag{5.210}$$

C_1, C_2 是比例系数；$\mathscr{E}$ 是漏端附近的电场强度；E_i 是热电子通过碰撞电离产生电子-空穴对所必须具有的最小能量；λ 可理解为热电子得到能量 $k_B T_e$（T_e 是热电子温度）所走过的平均路程。所以，$e^{-E_i/q\mathscr{E}\lambda}$（等于 $e^{-E_i/k_B T_e}$）表示 I_D 中具有足够能量引起碰撞电离的电子所占的比例，亦即 $I_b \propto I_D e^{-E_i/q\mathscr{E}\lambda}$。同理，$e^{-E_b/q\mathscr{E}\lambda}$ 是 I_D 中具有足够能量越过 Si/SiO_2 界面注入栅氧化层中的电子所占比例，亦即 $I_G \propto I_D e^{-E_b/q\mathscr{E}\lambda}$，$E_b$ 是 Si/SiO_2 界面的势垒高度。由上述二式消去 $q\mathscr{E}\lambda$，可得

$$\frac{I_G}{I_D} = C_2\left[\frac{I_b}{C_1 I_D}\right]^p \tag{5.211}$$

其中 $p = E_b/E_i$。所以可通过检测 I_b 来了解器件性能退化的情形。

关于热电子引起 Si/SiO_2 界面损伤或产生氧化物陷阱的机理，目前还是一个比较活跃的研究领域。已知在 Si 上热生长的 SiO_2 薄膜是非晶态的，但为了简单明确，这里还是采用了图 5.25 所示的假想模型来说明在热电子作用下界面陷阱产生及随时间

变化的规律。图 5.25(a)表示,在 Si/SiO_2 界面处,硅表面的原子只有 3 个最近邻,不能像体内的原子那样有 4 个最近邻可以形成共价键,即它有一个电子没有自旋配对,存在悬挂键。若材料中存在 H 原子(图中以"x"标记),这些悬挂键容易同 H 原子组成 Si-H 键,这时 Si 原子得到一个电子,成为带负电的离子。图 5.25(b)表示,在 A 处的 Si-H 键受到破坏,H 原子同 SiO_2 薄膜表面处的 O 原子结合,并破坏一个 Si-O 键,形成 Si-OH 键。这样,在 A 处留下一个电子陷阱,在 B 处则留下一个空穴陷阱。B 处的 Si 原子因缺少一个 O 近邻而有一个电子没有转移出去,所以是空穴陷阱。H 原子向 SiO_2 薄膜内部扩散,界面态浓度 N_{it}随时间增加。

如果 Si-H 键遭到破坏的原因是热电子,即如果界面陷阱由热电子引起,则其密度 N_{it}和 I_D 的关系可表示为

$$\frac{dN_{it}}{dt} = R\frac{I_D}{Z}e^{-Et/q\mathscr{E}\lambda} - BN_{it}n_{it}(0) \tag{5.212}$$

Z 是沟道宽度,热电子浓度正比于 I_D/Z,而使 Si-H 键离解的热电子数正比于$(I_D/Z)e^{-E_t/q\mathscr{E}\lambda}$,其中 $E_t \simeq (3.2+0.3)$eV,包括3.2eV的界面势垒激活能和 0.3eV 的 Si-H 键离解能;R 正比于 Si-H 键密度,后者高达 $10^{20}cm^{-3}$,故 R 可视为常数。$BN_{it}n_{it}(0)$表示界面的 Si 离子和 H 原子重新结合的速率,其中 $n_{it}(0)$是界面处的 H 原子浓度。另一方面,界面陷阱产生的速率也等于 H 原子离开界面的速率,即

$$\frac{dN_{it}}{dt} = D_H n_{it}(0)/L_H \tag{5.213}$$

D_H 和 L_H 分别表示 H 原子的扩散系数和扩散长度,故 D_H/L_H 为 H 原子的有效扩散速度。

由方程(5.213)求出 $n_{it}(0)$后代入方程(5.212),有

$$\frac{BL_H}{2D_H}n_{it}^2 + N_{it} = Rt\frac{I_D}{Z}e^{-ET/q\mathscr{E}\lambda} \tag{5.214}$$

这一方程表示,N_{it}增长的动力学过程十分类似热氧化过程,N_{it}很小时增长速率受反应过程限制,$N_{it} \propto t$;N_{it}很大时增长速率受扩散过程限制,$N_{it} \propto t^{1/2}$;t 是时间。在一般情形下,像热氧化情形一样,为

$$N_{it} = C_3\left[t\frac{I_D}{Z}e^{-E_t/q\mathscr{E}\lambda}\right]^n \tag{5.215}$$

C_3 近似为常数,n 的数值在 0.5 和 0.75 之间(理论上在 0.5 和 1 之间)。

MOSFET 的寿命 τ 定义为N_{it}达到某一标准值或阈值电压的变化量 ΔV_T($\Delta V_T \propto N_{it}$)达到某一失效标准时的时间。选取 $n=2/3$,$E_t/E_i=2.9$(其中 $E_t=3.7$eV 和 $E_i=1.3$eV 都来自实验数据),并利用 $I_b = C_1 I_D e^{-E_i/q\mathscr{E}\lambda}$消去 $q\mathscr{E}\lambda$,不难从式(5.215)得到

$$\tau = F\frac{Z}{I_D}\left(\frac{I_b}{I_D}\right)^{-2.9}(\Delta V_T)^{1.5} \tag{5.216}$$

F 依赖于栅绝缘层(氧化硅或氮化硅)的质量,如果不采用高质量的绝缘栅材料,热电子退化将限制器件向亚微米尺寸缩小。

另外,根据 I_b 和 I_D 之间的关系($I_b=(M-1)I_D$),不难从方程(5.216)得到

$$\tau \propto \frac{Z}{I_D}(M-1)^{-m} \tag{5.217}$$

式中 M 是碰撞电离倍增因子,$m=2.9$。

对于 $L>0.5\mu m$ 的 pMOST，器件性能退化并不是一个很严重的问题，这主要是由于 Si/SiO_2 界面处的空穴势垒要比电子势垒高得多，以及空穴电离率较小从而它引起的电子-空穴对产生率较低。但当器件进入深亚微米领域（$L<0.5\mu m$）以后，必须考虑 pMOST 的热电子退化效应。

热电子退化对 MOSFET 的可靠性构成威胁，必须尽可能避免，因此漏区结构的设计受到了很大的关注。所有结构中的基本原则是一样的，就是让漏区承受一部分电压。一种典型结构是所谓轻掺杂漏（LDD）结构，即在有效沟道和重掺杂的漏接触之间增加一个高阻区。

5.2.7 CMOS（互补型 MOSFET）结构

CMOS 技术一直是硅微电子技术的主流。为此，下面简述 CMOS 器件结构的基本特点。以 CMOS 为基本单元的 CMOS 集成电路具有功耗低、抗干扰能力强和速度快的优点，在一般逻辑电路、大规模存储器以及微处理器等领域都得到了广泛的应用。

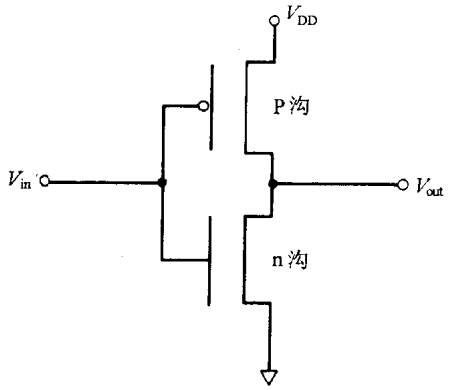


图 5.26 CMOS 反相器

图 5.26 表示一个 CMOS 反相器。上部的 pMOST 的栅极与下部的 nMOST 的栅极相连，两个晶体管都是增强型，pMOST 的阈值电压 $V_{Tp}<0$，nMOST 的阈值电压 $V_{Tn}>0$。当输入电压 V_{in} 为 0 或一个小的正值（$<V_{Tn}$）时，pMOST 导通（pMOST 的栅压 $V_{GSp}=V_{in}-V_{DD}$，此时比 V_{Tp} 更负），而 nMOST 截止，因此输出电压 V_{out}（等于 V_{DSn}）为高电平而十分接近电源电压 V_{DD}（逻辑 1）。当输入电压为 V_{DD} 时，pMOST 因其栅压 $V_{GSp}=0$ 而截止，而 nMOST（栅压 $V_{GSn}=V_{in}>V_{Tn}$）导通，此时输出电压 V_{out} 为低电平而接近零（逻辑 0）。不论哪种逻辑状态，串联在 V_{DD} 和地之间的这两个晶体管中，总有一个处于非导通状态，所以稳态电流只是很小的泄漏电流，只是在开关过程中两个管子都处在导通状态时才有显著的电流流过 CMOS 反相器，故平均功耗很小，在毫微瓦数量级。当芯片的元件数增加时，功耗成为主要的限制因素，CMOS 电路的低功耗是十分引人注目的特点。

图 5.27 表示 nMOST 和 pMOST 完全对称的 CMOS 反相器的准静态传输特性（V_{in} 缓慢变化时 V_{out} 同 V_{in} 的关系）。所谓完全对称，指两个管子的阈值电压相等和导电因子 $\left(\text{定义为}\frac{Z}{2L}\mu C_{ox}\right)$ 相同，即 $V_{Tn}=|V_{Tp}|=V_T$ 和 $\beta_n=\beta_p=\beta$。对于这种反相器，两种逻辑状态之间的转换电压 V_{in}^*（这时两个管子都处在饱和区，由电流连续条件有 $\beta_p(V_{in}^*-V_{DD}-V_{Tp})^2=\beta_n(V_{in}^*-V_{Tn})^2$）为

$$V_{in}^* = V_{DD}/2 \tag{5.218}$$

V_{in}^* 也称反相器的阈值电压。对应 V_{in}^*，输出电压以 $V_{in}^*-V_{Tp}$ 突变到 $V_{in}^*-V_{Tn}$，传输特性成为一段竖直线，其高度为 $|V_{Tp}|+V_{Tn}\leqslant V_{DD}$，对于完全对称的情形，显然有

$$V_T \leqslant V_{DD}/2 \tag{5.219}$$

图 5.27 中画出了 $V_T/V_{DD}=0.5, 0.25, 0.1$ 三种情形下的静态传输特性。$V_T/V_{DD}=0.5$ 的特性曲线最陡峭，随输入电压从低电平向高电平增加，$V_{in}=V_{DD}/2$ 时发生开关过程，pMOST 立即关

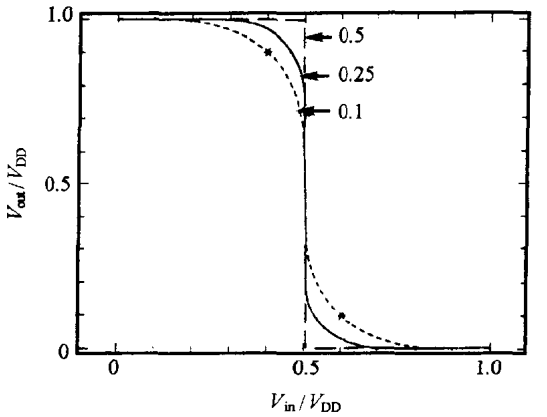


图 5.27　$V_T/V_{DD}=0.5, 0.25, 0.1$ 三种情形下 CMOS 反相器的准静态传输特性(＊号表示单位增益点)

断,nMOST则勉强导通,结果因为导通电流很小(导通电阻 R 很大),输出结点电容放电十分缓慢。这样,在动态情形下,即使输出结点的电容很小,V_{out}也是随 V_{in}的继续增加而缓慢下降。所以,对于合理的电路速度,V_T/V_{DD}应显著小于 0.5,即 V_{DD}应显著超过 $2V_T$,而且超过越多,电路的速度越快。另一方面是考虑噪声容限。CMOS 反相器的噪声容限由传输特性曲线上两个单位增益($|dV_{out}/dV_{in}|=1$)点的位置确定,如图 5.27 中 $V_T/V_{DD}=0.1$ 的传输曲线上标有星号"＊"的点所示。当输入电压在单位增益点之间时,通过反相器的信号噪声会被放大,即输入电压落入这一区间时会引起输出误差。这一区间越窄越好,由图 5.27 不难看出,这要求 V_{DD}尽量接近 $2V_T$。所以,在电路速度和噪声容限之间折衷考虑,一个较好的选择是

$$V_{DD} = 4V_T \tag{5.220}$$

顺便指出,对于 CMOS 电路的另一个重要的基本单元,CMOS 传输门,为了减小 RC 时间延迟,V_{DD}应选择比 V_T 高很多。分析表明,一个合适的结果也是 $V_{DD}\simeq 4V_T$。

CMOS 工艺要求在同一衬底上制造出 n 沟和 p 沟晶体管,因此必须在原始衬底上形成一个掺杂类型相反的区域,该区域被称之为"阱"。例如,在 n 阱 CMOS 工艺中,利用 p 型衬底制作 nMOST,n 阱制作 pMOST。除 n 阱工艺外,还有 p 阱工艺和双阱工艺。但最常用的是 n 阱工艺,这主要是因为 n 沟 MOSFET 的设计技术比较成熟,容易将该技术移植到 CMOS 电路的设计中。此外,对于短沟道的 nMOST,热电子效应非常严重,而与制作在 p 阱中的 nMOST 相比,制作在 p 型衬底上的 nMOST 更容易实现与衬底的低阻连接。

图 5.28 给出 n 阱 CMOS 器件结构的截面图。从中可以发现,CMOS 结构在水平和垂直方向上形成了两个寄生的双极晶体管,即 p^+源、n 阱和 p 型衬底在垂直方向上形成的 p-n-p 晶体管,n 阱、p 型衬底和 n^+源在水平方向上形成的横向 n-p-n 晶体管,每个晶体管的基极都由另一个晶体管的集电极驱动,因此形成了一个反馈回路,该 p-n-p-n 反馈回路实际上是一个可控硅整流器(SCR),它的增益等于 n-p-n 和 p-n-p 晶体管的共发射极增益 β_{npn}和 β_{pnp}的乘积。当回路增益大于 1 时,SCR 变成低阻态,有很大的电流通过,这种现象称为闩锁效应。闩锁效应极易破坏芯片,使电路失效。在某些不稳定因素的触发下,例如瞬时电流或电离辐射等在阱和衬底中引起横向电流,双极晶体管的发射极-基极结正偏,激活寄生的可控硅开关形成闩锁。

要防止闩锁效应，寄生晶体管的电流增益 β_{npn}和 β_{pnp}一定要小，阱和衬底的电阻 R_{well}和 R_{sub}一定要小。这些要求可以通过采用适当的结构和工艺方案来实现，比一般 CMOS 工艺更重要的是所谓“在绝缘膜上生长硅”的 SOI CMOS 工艺。这些工艺不属于本书的范围，我们不讨论。

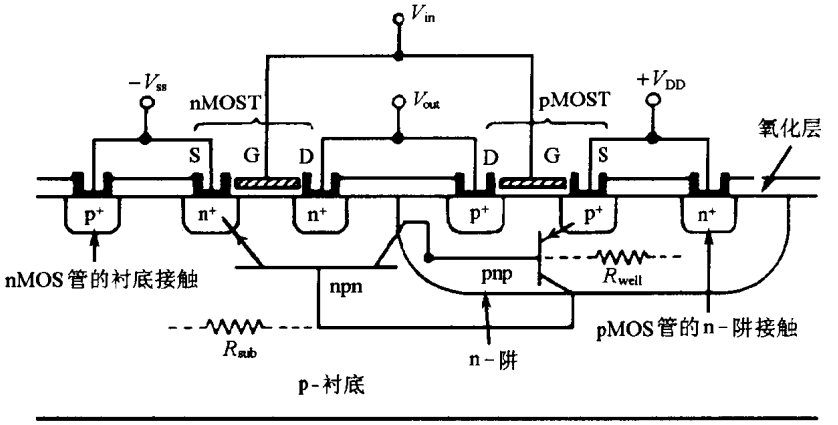


图 5.28　在 n 阱 CMOS 工艺中，n 沟和 p 沟 MOSFET 的截面图

5.3　短沟道 MOSFET

为了提高工作速度和单位芯片面积的集成度，都需要把 MOSFET 做得尽可能小。目前，MOSFET 的沟道长度已进入 0.1～0.5μm 范围，芯片上集成的器件数在 10^8～10^9 量级。然而沟道长度(栅长)很小时，如果不在器件结构上认真考虑，就会出现我们在 5.2.6 节中所讨论的各种短沟道效应，器件不能正常工作。为此，本节专门就 MOSFET 小型化的理论与实践问题进行一定的讨论。

5.3.1　器件小型化的规则

减小沟道长度而又能保持原来长沟道特性的一个简单方法，是将器件的所有尺寸和电压都按同一比例因子$\frac{1}{\alpha}$($\alpha>1$)缩小，使器件内电场的分布和强度与长沟道 MOSFET 的情形相同。具体来说，L，Z，d_{ox}，X_j，V_{GS}，V_{DS}和 V_{BS}都缩小到$\frac{1}{\alpha}$倍。为了使耗尽区厚度也按$\frac{1}{\alpha}$缩小，衬底掺杂浓度应增加到 α 倍。器件的阈值电压因为 ϕ_{ms}和 ϕ_F 不按$\frac{1}{\alpha}$缩小，则必须用离子注入沟道区表面进行调节，使之也按$\frac{1}{\alpha}$缩小。这种按比例缩小的方法通常称为恒定电场规则(CE 规则)。按照 CE 规则，一些物理量的换算关系如下：

$$C'_{ox}=\frac{\varepsilon_{ox}}{d'_{ox}}=\alpha C_{ox}[\mathrm{F/cm^2}] \tag{5.221}$$

$$(C_{ox}A)'=C'_{ox}L'Z'=\frac{C_{ox}A}{\alpha}[\mathrm{F}] \tag{5.222}$$

$$I'_D=\frac{1}{\alpha}I_D[\mathrm{A}] \tag{5.223}$$

$$J_D' = \alpha J_D[\mathrm{A/cm^2}] \tag{5.224}$$

$$f_T' = \frac{g_m'}{2\pi(C_{ox}A)'} = \alpha f_T[\mathrm{Hz}] \tag{5.225}$$

功耗 P_{dc}和延迟时间 t_{pd}分别换算为

$$P_{dc}' = I_D' V_{DS}' = \frac{1}{\alpha^2} P_{dc}[\mathrm{W}] \tag{5.226}$$

$$t_{pd}' = \frac{1}{\alpha} t_{pd}[\mathrm{S}] \tag{5.227}$$

t_{pd}主要决定于对器件电容充放电的时间，即 $t_{pd} \propto CV/I$。功率-延时乘积是数字电路的一个质量指数，将有

$$P_{dc}' t_{pd}' = \frac{1}{\alpha^3} p_{dc} t_{pd} \tag{5.228}$$

亚阈值摆幅 S 正比于$(1 + C_D/C_{ox})$，而 C_D 和 C_{ox}按同一比例$\frac{1}{\alpha}$变化，所以有

$$S' \simeq S \tag{5.229}$$

即亚阈值斜率不能按比例缩小。所以，按 CE 规则缩小的器件中，亚阈值电流将增大，而栅压大于阈值时的电流将减小，从而使器件的开关特性变差。对于数字化电路而言，这是人们所不希望的。此外，由于实际应用和标准化等原因，电源电压也不能按 CE 规则变化。

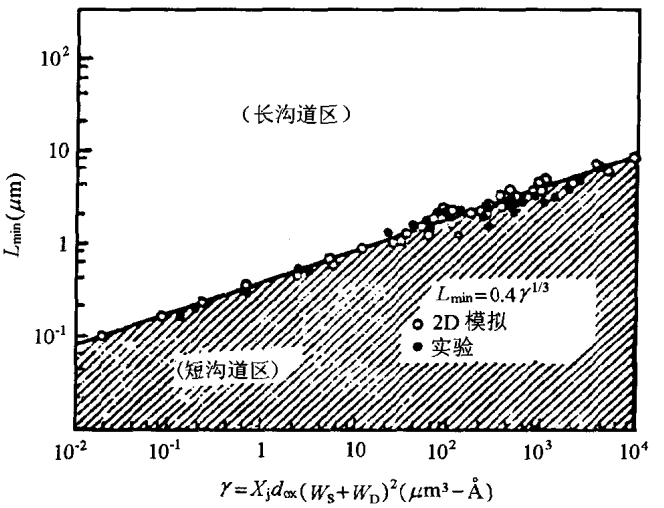


图 5.29　最小沟道长度 L_{min}和 γ 的关系

目前特受关注的缩小 MOS 器件尺寸的方法，是在努力减小尺寸的同时保持长沟道器件的亚阈值区特性，这种特性是不出现短沟道效应的重要标志。根据广泛的实验研究和二维数值模拟，发现小型化中保持长沟道器件特性的最小沟道长度满足下述经验关系：

$$L_{min} \simeq 0.4[X_j d_{ox}(W_S + W_D)^2]^{1/3} = 0.4\gamma^{1/3} \tag{5.230}$$

图 5.29 表示 L_{min}和 $\gamma = X_j d_{ox}(W_S + W_D)^2$ 的关系。其中，X_j 是重掺杂源(漏)区的结深；d_{ox}是栅氧化层厚度；W_S 和 W_D 分别是源结和漏结的耗尽层宽度，对于均匀掺杂的突变结，它们的表示式为

$$W_S = \sqrt{\frac{2\epsilon_s}{qN_A}(V_{bi} + V_{SB})} \tag{5.231}$$

$$W_D = \sqrt{\frac{2\epsilon_s}{qN_A}(V_{bi} + V_{SB} + V_{DS})} \tag{5.232}$$

V_{bi}是内建电压，V_{SB}是源-衬偏压，N_A是衬底掺杂浓度。X_j，W_S，W_D的单位用微米，d_{ox}用Å，L_{min}的单位是微米。

方程(5.230)表示MOSFET的沟道长度不能小于L_{min}。按照这一方程，小型化要求浅结(X_j)、薄栅氧化层(d_{ox})、低电压或高掺杂使($W_S + W_D$)减小。我们将围绕这些变量组织下面的讨论，分析如何设定它们的最低极限。

5.3.2 结深 X_j

首先，结深减小会使结的横截面积(等于结深和沟道宽度的乘积)减小，寄生的串联电阻R_S和R_D增加，结果如5.2.5节所分析指出的，器件的跨导和电导下降，驱动电流减小。下面进一步分析影响R_S和R_D的因素。

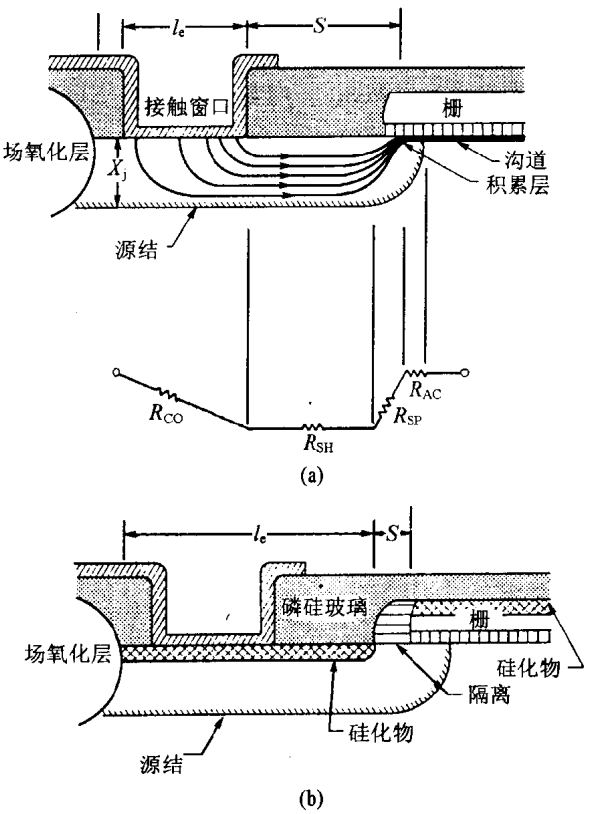


图 5.30 (a) MOSFET 源(漏)区串联电阻的各个分量；(b) 硅化物接触

图5.30(a)给出了源区电流分布的示意图，可以看出，寄生串联电阻R_S可以分成四部分：接触电阻R_{CO}，重掺杂源扩散区的薄层电阻R_{SH}，源区沟道一侧的端点处由于电流集中而产生的扩展电阻R_{SP}，扩展电阻区和实际的沟道端点之间的积累层电阻R_{AC}。薄层电阻可简单地由下式得到：

$$R_{SH} = \frac{\rho S}{X_j Z} \tag{5.233}$$

ρ 是扩散区的平均电阻率，Z 是沟道宽度，S 是电流分层流动区域的长度，X_j 是扩散区的厚度（即结深）。扩展电阻 R_{SP}是由于 MOSFET 的沟道电流向外呈放射状地扩展产生的，开始时电流被限制在栅-源交叠区的表面积累层内，所以通常把 R_{SP}和积累层电阻 R_{AC}一起考虑。假定源区掺杂均匀，在一级近似条件下有

$$R_{SP} + R_{AC} = \left(\frac{2\rho}{\pi Z}\right) \cdot \ln\left(H \frac{X_j}{t_{AC}}\right) \tag{5.234}$$

其中 t_{AC}是表面积累层的厚度，H 是一个大小在 0.37～0.9 之间的因子。由于 H 是方程中的对数项，而且比例项 X_j/t_{AC}的数值很大，因此 H 起的作用较小。在接触区内，扩散区的电压降使电流在接触窗口的前端附近集中，形成接触电阻 R_{CO}。采用金属和扩散区之间界面的传输线模型，对于接触边长度为 l_c 的矩形，其接触电阻 R_{CO}可表示为

$$R_{CO} = \frac{1}{Z}\sqrt{\frac{\rho\rho_c}{X_j}}\coth\left(l_c\sqrt{\frac{\rho}{\rho_c X_j}}\right) \tag{5.235}$$

ρ_c 是金属和源区之间的接触电阻率。如果 l_c 满足下述不等式：

$$l_c \gtrsim 1.5\sqrt{\frac{\rho_c X_j}{\rho}} \tag{5.236}$$

则式(5.235)中的双曲余切函数不会超出其渐近值 1 的 10%。这样，对于较大的 l_c，接触电阻将有极小值：

$$R_{CO} \simeq \frac{1}{Z}\sqrt{\frac{\rho\rho_c}{X_j}} \tag{5.237}$$

为了降低源(漏)区串联电阻，降低接触电阻率 ρ_c 至关重要。为此，应努力提高欧姆接触技术。

现在经常采用在露出的硅表面上形成自对准硅化物(Self-aligned silicide，简写为 Salicide)技术，使串联电阻大大降低，如图 5.30(b)所示。硅化物的电阻率相当低(例如 $TiSi_2$ 和 $TaSi_2$ 等硅化物的电阻率低至 50$\mu\Omega$·cm 以下)，工艺上也和集成电路兼容，随着器件尺寸的缩小，它们是越来越重要的金属化材料。由该图可以看出，接触窗口的金属层通过硅化物延伸，直至由侧墙氧化物(厚度 S)同硅化物栅隔离。由于 S 很小，没有薄层电阻区(即 $R_{SH}=0$)，串联电阻 $R_S(R_D)$变为 $R_{CO}+R_{SP}+R_{AC}$。

结深减小还容易使表面损伤落入结的耗尽层内，或使硅化物/硅界面靠近耗尽层，引起器件严重漏电。前者是因为耗尽层内复合中心增加；后者是因为形成硅化物时要消耗一定重掺杂的硅，从而耗尽层的厚度增加。这两点是形成浅结的主要障碍。通过改进工艺，降低表面损伤，结深可以做到比较小。在目前的工艺水平下，$X_j \simeq 500$Å 已经能够实现，将来还有希望做得更小一些。

5.3.3 栅氧化层厚度 d_{ox}

沟道载流子(对 nMOST 为电子)注入栅氧化层，可以通过前面叙述过的热电子效应，即高能电子越过 Si/SiO_2 界面势垒；也可以通过隧穿效应，包括直接隧穿和电场辅助隧穿(或称 Fowler-Nordheim 隧穿，F-N 隧穿)。图 5.31 示出了上述三种过程。隧穿过程和薄栅氧化层直接关联，或者和与之相联系的十分高的栅氧化层纵向电场 $\mathscr{E}_{ox}$有关，所以通过对它的讨论可以

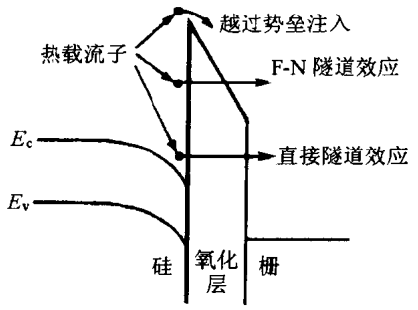


图 5.31 沟道载流子注入栅氧化层的三种过程

给出减小栅氧化层厚度的限制条件。

对于一定栅压，随着栅氧化层厚度 d_{ox}的减小，栅氧化层电场 $\mathscr{E}_{ox}$增大。在 $\mathscr{E}_{ox}$足够高(>6MV/cm)时，沟道电子的运动方向受到强烈影响，一些电子(所谓"幸运"电子)通过隧道效应进入栅氧化层，引起栅极电流，即 Fowler-Nordheim 电流。这种高场栅极电流在氧化层及其界面产生陷阱或电荷，并因电荷积累作用引起栅氧化层击穿，栅和衬底永久短路。

如果栅氧化层电场均匀(实际上由于氧化层内存在电荷，很难说电场是均匀的)，则 *F-N* 电流密度可表示为

$$J = A\mathscr{E}_{ox}^{2} e^{-B/\mathscr{E}_{ox}} \tag{5.238}$$

其中 $A = 1.25\times10^{-6}\text{A/V}^2$，$B = 2.34\text{MV/cm}$。

通常由氧化层漏电流应小于源(漏)结漏电流的条件来估计可以容许的最大栅氧化层电场 $\mathscr{E}_{max}$。例如，如果规定一个约等于 10^{-10}A/cm^2 的电流值，则由式(5.238)有

$$\mathscr{E}_{max} \simeq 5.8\text{MV/cm} \tag{5.239}$$

为了 MOSFET 正常工作，栅氧化层电场 $\mathscr{E}_{ox}$应低于这一数值。

$\mathscr{E}_{max}$给出了栅氧化层的最小允许厚度 d_{min}，为

$$d_{min} = \frac{V_{GS} - V_{FB} - \psi_s}{\mathscr{E}_{max}} \tag{5.240}$$

ψ_s 为半导体的表面势，即半导体表面层上的电压降。上式已假定了有效界面电荷 $Q_0 = 0$。V_{GS}可能因 MOSFET 的应用而异，但电路中通常使用统一的工作电压，所以在这里代之以电源电压 V_{DD}，而将上式改写成

$$d_{min} = \frac{V_{DD} - V_{FB} - \psi_s}{\mathscr{E}_{max}} \tag{5.241}$$

另一方面，当氧化层十分薄($d_{ox}\lesssim30\text{Å}$)时，载流子有显著的几率隧穿氧化层，在栅和衬底之间直接流动。对于这种直接隧道效应，由于与氧化层质量和界面形态密切相关，很难从计算隧穿电流来得到受其限制的最小栅氧化层厚度，而只能根据实验结果把它设定为

$$d_{min} = 30\text{Å} \tag{5.242}$$

5.3.4 耗尽层宽度 $W_{S,D}$

为了减小耗尽层宽度 W_S 和 W_D，必须提高沟道区掺杂浓度或降低偏压。

1. *沟道掺杂浓度和阈值电压*

作为粗略近似，考虑长沟道和均匀掺杂，这时阈值电压可表示为

$$V_T = V_{FB} + 2\phi_F - Q_s/C_{ox} \tag{5.243}$$

它包含平带电压 V_{FB}、强反型时的半导体表面电压(表面势)和相应的表面层电荷 Q_s 在栅氧化层上产生的电压，其中

$$\phi_F = \frac{1}{\beta}\ln\frac{N_A}{n_i} \tag{5.244}$$

$$Q_s = -qN_A\sqrt{2}L_D(2\beta\phi_F - 1)^{1/2} \tag{5.245}$$

式(5.245)由式(5.31)令 $\psi_s = 2\phi_F$ 得到，括号中的因子 1 通常可以忽略。由于栅氧化层厚度 d_{ox}(即 C_{ox})已由栅氧化层最大允许电场 $\mathscr{E}_{max}$或最小允许厚度 d_{min}限定，V_T 实际上只决定于沟道区掺杂浓度 N_A。换句话说，N_A 应由 V_T 确定；而 V_T 应当和电源电压 V_{DD}协调，使器件表现出最佳性能。V_T 和 V_{DD}的关系可能因 MOSFET 的应用而异，考虑到它主要用于 CMOS 集成电路，故选择(参见 5.2.7 节)

$$V_{DD} = 4V_T \tag{5.246}$$

2. *沿沟道方向(横向)电场和耗尽层宽度*

根据两区模型，漏端高场区(速度饱和区或夹断区)内的横向电场可表示为

$$\mathscr{E}_m = \frac{V_{DD} - V'_{Dsat}}{l} \tag{5.247}$$

l 是速度饱和区的特征长度，可近似看作耗尽层宽度，并且用经验公式(5.208)来确定。V'_{Dsat} 是饱和电压，表示为[参见(5.190)式]

$$V'_{Dsat} = \frac{(V_{GS} - V_T)L\mathscr{E}_c}{(V_{GS} - V_T) + L\mathscr{E}_c} \tag{5.248}$$

通常，比起漏击穿来，热电子退化是对最大横向电场的更严厉的限制(发生漏击穿的电场估计约 0.6MV/cm，而热电子退化时约 0.2MV/cm)，限制条件可设定为

$$\mathscr{E}_m < 0.2\text{MV/cm} \tag{5.249}$$

由式(5.247)可知，设定 $\mathscr{E}_m$ 也就规定了 V_{DD}的上限；而后者又是随沟道长度(栅长)L 而变的，从式(5.248)可以看出这一点。

$\mathscr{E}_m$ 受热电子退化限制，器件寿命 τ 描述了热电子退化条件，从式(5.217)可得

$$\tau = BZ\frac{(I_b/I_D)^{-m}}{I_D} \tag{5.250}$$

I_b 是衬底电流，代替表示式(5.209)，这里用半经验公式：

$$I_b = \gamma I_D\left(\frac{\mathscr{E}_m}{\mathscr{E}_i}\right)e^{-\mathscr{E}_i/\mathscr{E}_m} \tag{5.251}$$

其中 $\gamma \simeq 3\times10^{10}\text{cm}^{-2}$；$\mathscr{E}_i \simeq 2\times10^6\text{V/cm}$；$m$ 与氧化层电场有关，对于 nMOST，在前面的讨论中得到 $m \simeq 3$；B 依赖于热电子注入、栅氧化层质量和对寿命的选择标准(通常选择栅压和漏压一定时漏极电流 I_D 因热电子损伤改变 10%)。式(5.251)说明 I_b/I_D 主要是横向电场 $\mathscr{E}_m$ 的函数。

热电子只是当 $\mathscr{E}_m$ 高到足以引起速度饱和时才是值得注意的。饱和时漏极电流为

$$I_D = ZC_{ox}v_s\frac{(V_{GS} - V_T)^2}{(V_{GS} - V_T) + L\mathscr{E}_c} \tag{5.252}$$

v_s 是载流子的饱和速度，对 Si 中的电子有 $v_s \sim 10^7\text{cm/s}$。将上式代入式(5.250)可得

$$\frac{v_s\tau}{B} = \left(\frac{I_D}{I_b}\right)^m\frac{1}{C_{ox}(V_{GS} - V_T)}\left(1 + \frac{L\mathscr{E}_c}{V_{GS} - V_T}\right) \tag{2.253}$$

方程(5.253)表示，MOSFET 的寿命 τ 可以看作是栅氧化层纵向电场 $\mathscr{E}_{ox}$(通过 m 及 $C_{ox}\times(V_{GS} - V_T)$)和漏端横向电场 $\mathscr{E}_m$(通过 I_b/I_D)的函数。器件的寿命标准决定了可允许的

最大 $\mathscr{E}_m$($\mathscr{E}_{ox}$的最大允许值已在5.3.3节讨论),作为大致估计,设其为0.2MV/cm。

5.3.5 最短沟道长度 L_{min}

利用方程(5.230)来确定MOSFET最短沟道长度的一条途径,是从用 $\mathscr{E}_{ox}$的约束条件确定沟道掺杂浓度 $N(N_A)$出发。当栅氧化层电场为 $\mathscr{E}_{max}$时,半导体的表面势 ψ_s 在一维近似下由高斯定理给出(参见方程(5.31)):

$$\varepsilon_{ox}\mathscr{E}_{max} = qN\sqrt{2}L_D\left[e^{\beta(\psi_s-2\phi_F)} + \beta\psi_s - 1\right]^{1/2} \tag{5.254}$$

其中

$$L_D = \sqrt{\frac{\varepsilon_s}{\beta qN}},\quad \beta = \frac{q}{k_B T}$$

上述方程的右方为半导体表面层电荷(面密度)Q_s 之值,显然在写出这一方程时忽略了有效界面电荷 Q_0。这一方程可以改写为

$$\beta\psi_s = 2\beta\phi_F + \ln\left[\frac{1}{2}\left(\frac{\varepsilon_{ox}}{\varepsilon_s}\right)^2(\beta\mathscr{E}_{max}L_D)^2 + 1 - \beta\psi_s\right] \tag{5.255}$$

方程(5.255)可以用 $\psi_s = 2\phi_F$ 作为初值用迭代法求解。

阈值电压可近似为(参见式(5.243))

$$V_T = V_{FB} + 2\phi_F + \sqrt{2}\left(\frac{\varepsilon_s}{\varepsilon_{ox}}\right)\left(\frac{d_{ox}}{L_D}\right)\frac{k_B T}{q}(2\beta\phi_F)^{1/2} \tag{5.256}$$

设 n^+ 硅栅,忽略有效界面电荷 Q_0 时,平带电压 V_{FB}为

$$V_{FB} = -\phi_G - \phi_F \tag{5.257}$$

ϕ_G 为半值禁带电压,300K时有 $\phi_G = 0.56$V。

取栅偏压为 $V_{DD} = 4V_T$,则可以通过下式求栅氧化层厚度 d_{ox}:

$$\mathscr{E}_{max} = \frac{4V_T - V_{FB} - \psi_s}{d_{ox}} \tag{5.258}$$

将式(5.255),(5.256)和(5.257)代入上式并加整理,得

$$\frac{d_{ox}}{L_D} = \beta\frac{-3\phi_G + 5\phi_F - \psi_s}{\beta\mathscr{E}_{max}L_D - 4\sqrt{2}(\varepsilon_s/\varepsilon_{ox})(2\beta\phi_F)^{1/2}} \tag{5.259}$$

这样,对于设定的 $\mathscr{E}_{max}$,栅氧化层厚度 d_{ox}将是沟道区掺杂浓度 N 的函数;即由选定 N,可得到 d_{ox}。

如前所述,耗尽区宽度($W_S + W_D$)可以由速度饱和区的特征长度 l 近似,且

$$l \simeq \begin{cases} 0.22 d_{ox}^{1/3} X_j^{1/3} & (d_{ox} > 150\text{Å}) \\ 0.017 d_{ox}^{1/8} X_j^{1/3} L^{1/5} & (d_{ox} < 150\text{Å}\ L < 0.5\mu\text{m}) \end{cases} \tag{5.260}$$

d_{ox}, X_j 和 L 的单位都用厘米(cm)。

现在,可以在给定结深 X_j 和掺杂浓度 N 的条件下,通过方程(5.230)来计算 L。例如,对 $X_j = 500$Å 和 $N = 1.23\times10^{18}\text{cm}^{-3}$($d_{ox} = 35$Å),有 $L \simeq 0.1\mu$m。另一方面,根据方程(5.256),相应地得到 $V_T \simeq 0.53$V($V_{DD} \simeq 2.1$V);用 V_{DD}的数据,根据方程(5.247)和(5.248)可计算 $\mathscr{E}_m$,结果是0.38MV/cm,它不满足 $\mathscr{E}_m < 0.2$MV/cm 的条件。所以,从防止热电子退化的角度,必须改善氧化层抗热电子损伤的能力,或者采用像轻掺杂漏(LDD)之类的特殊结构。LDD 适

用于大尺寸的器件，对小尺寸器件来说在工艺上难以控制。

降低使用电压是防止热电子损伤的有效途径。为此，可以选定 X_j 和 d_{ox}，根据方程(5.230)来计算 L；并在 $\mathscr{E}_m=0.2$MV/cm 的极限条件下求出 V_T(或 V_{DD})，确定沟道区所需的掺杂浓度 N。例如，$X_j=500$Å 和 $d_{ox}=40$Å 时，得到 $L\simeq 0.13\mu$m 和 $N\simeq 5.5\times 10^{17}\text{cm}^{-3}$。由于使用了较低的电压，栅氧化层电场 $\mathscr{E}_{ox}$也会有所降低。对于上述例子，$V_{DD}=1.4$V($V_T=0.35$V)，将有 $\mathscr{E}_{ox}\simeq 3.8$MV/cm 而低于极限值 $\mathscr{E}_{max}=5.8$MV/cm。应当指出，结深 X_j 并不是决定沟道长度 L 的关键参数。例如，上述 $L=0.13\mu$m 的器件在 $X_j=800$Å 和 $V_T=0.36$V 时也可以得到，但是必须采用 $d_{ox}=30$Å 的栅氧化层。换句话说，根据最小沟道长度的经验公式缩小器件尺寸的方法具有较大的灵活性，它允许在保持 $\gamma=[X_j d_{ox}(W_S+W_D)^2]$一定的条件下独立调整器件的参数，这样使器件容易制造，或者使器件的某些性能得到优化。

5.4 SOI MOSFET

这是用 SOI(Silicon on Insulator)技术制造的器件，其中生长在绝缘体上的半导体薄膜构成器件的有源部分。薄膜 SOI MOSFET 能有效地抑制小尺寸效应，在 VLSI 中具有明显的优势；同时 SOI 层能够堆砌，具有形成三维集成电路的潜力。近年来，SOI 技术发展很快，并日趋成熟，这些技术包括在硅中形成多孔硅(FIPOS 技术)和利用离子束合成在硅中形成绝缘层(SIMOX, SIMNI 和 SIMON)等，其中用得较为普遍的是 SIMOX(Sepatation by IMplanted Oxygen，注氧隔离技术)。本节讨论 SOI MOSFET 的基本特性。

5.4.1 概述

图 5.32 示意画出了 SOI MOSFET 的结构，t_{oxf}是正面栅氧化层厚度，t_{oxb}是背面栅氧化层厚度，t_s 是硅膜厚度。V_{Gf}，V_{Gb}分别为正、背栅偏压。

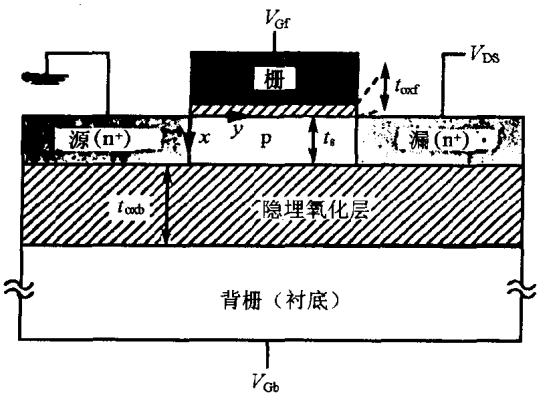


图 5.32 SOIn 沟 MOSFET 的剖面图

根据 Si 膜厚度 t_s 和表面耗尽层最大厚度 X_d 的比较，SOI MOSFET 分为三种不同的类型：厚膜器件、薄膜器件和中等膜厚器件。X_d 一般表示为 $\sqrt{\frac{2\varepsilon_s}{qN_A}(2\phi_F)}$，$\phi_F$ 是 Si 材料的费米势，表示为 $\frac{k_B T}{q}\ln(N_A/n_i)$，$N_A$ 为 Si 材料的掺杂浓度。

对于厚膜器件，$t_s > 2X_d$，正、背面的表面耗尽层之间不相互影响，在它们之间存在一层中性区，如果将这一中性区接地，则厚膜器件和通常的 MOSFET(体硅器件)特性上几乎完全相同；如果中性区不接地而处于浮空状态，则漏源之间形成一个 n-p-n 寄生晶体管，影响器件特性。

对于薄膜器件，$t_s < X_d$，如果背栅加相当大的负偏压或正偏压，则器件的背面积累或反型，Si 膜并不耗尽。在各类 SOI 器件中，背面处于耗尽状态的薄膜器件最富有吸引力，这类器件具有优异的短沟道特性和近似理想的亚阈斜率等优点，在 VLSI 中具有明显的优势。

"中等膜厚"的器件的 Si 膜厚度满足 $X_d < t_s < 2X_d$，这类器件在不同的背栅压 V_{Gb} 下，或者呈现薄膜器件特性，或者呈现厚膜器件特性。

下面讨论 n 沟 SOI 薄膜场效应晶体管的特性。

5.4.2 阈值电压

硅膜中电势作为深度 x 的函数 $\psi(x)$ 在耗尽近似下满足下述泊松方程：

$$\frac{d^2\psi(x)}{dx^2} = \frac{qN_A}{\varepsilon_s} \tag{5.261}$$

把方程积分二次以后得到

$$\psi(x) = \frac{qN_A}{2\varepsilon_s}x^2 + \left(\frac{\psi_{sb} - \psi_{sf}}{ts} - \frac{qN_A t_s}{2\varepsilon_s}\right)x + \psi_{sf} \tag{5.262}$$

ψ_{sf}，ψ_{sb} 分别表示正、背表面处的电势。

硅膜中的电场分布通过 $\mathscr{E}(x) = -d\psi(x)/dx$ 由下式给出：

$$\mathscr{E}(x) = -\frac{qN_A}{\varepsilon_s}x + \frac{\psi_{sf} - \psi_{sb}}{t_s} + \frac{qN_A t_s}{2\varepsilon_s} \tag{5.263}$$

在上式中分别令 $x = 0$ 和 $x = t_s$，即得正、背界面处的半导体表面电场，分别为

$$\mathscr{E}_{sf} = \frac{\psi_{sf} - \psi_{sb}}{t_s} + \frac{qN_A t_s}{2\varepsilon_s} \tag{5.264}$$

$$\mathscr{E}_{sb} = -\frac{qN_A}{\varepsilon_s}t_s + \mathscr{E}_{sf} \tag{5.265}$$

根据电荷平衡原理，可以得到正栅氧化层上的电压：

$$\psi_{oxf} = \frac{\varepsilon_s\mathscr{E}_{sf} - Q_{0f} - Q_{if}}{C_{oxf}} \tag{5.266}$$

Q_{0f} 是正界面有效电荷密度，Q_{if} 是正面沟道电荷面密度($Q_{if} < 0$)，C_{oxf} 是正栅氧化层单位面积电容。同理，背栅氧化层上的电压为

$$\psi_{oxb} = \frac{-\varepsilon_s\mathscr{E}_{sb} - Q_{0b} - Q_{sb}}{C_{oxb}} \tag{5.267}$$

Q_{0b} 是背界面有效电荷密度，Q_{sb} 是背面处于反型($Q_{sb} < 0$ 或积累 $Q_{sb} > 0$)状态时的背沟道电荷密度，C_{oxb} 是背面栅氧化层单位面积电容。

正、背栅压可分别表示为

$$V_{Gf} = \psi_{sf} + \psi_{oxf} + \phi_{msf} \tag{5.268}$$

$$V_{Gb} = \psi_{sb} + \psi_{oxb} + \phi_{msb} \tag{5.269}$$

ϕ_{msf}，ϕ_{msb} 分别为正、背栅功函数差。

由式(5.264),(5.266)和(5.268)得到

$$V_{Gf} = V_{FBf} + (1 + C_s/C_{oxf})\psi_{sf} - \frac{C_s}{C_{oxf}}\psi_{sb} - \frac{\frac{1}{2}Q_B + Q_{if}}{C_{oxf}} \tag{5.270}$$

其中,$C_s = \varepsilon_s/t_s$ 是硅膜电容,$Q_B = -qN_A t_s$ 是硅膜耗尽电荷,$V_{FBf} = \phi_{msf} - Q_{0f}/C_{oxf}$是正栅平带电压。

由式(5.265),(5.267)和(5.269),背栅压可表示为

$$V_{Gb} = V_{FBb} - C_s/C_{oxb}\psi_{sf} + (1 + C_s/C_{oxb})\psi_{sf} - \frac{\frac{1}{2}Q_B + Q_{sb}}{C_{oxb}} \tag{5.271}$$

$V_{FBb} = \phi_{msb} - Q_{0b}/C_{oxb}$是背栅平带电压。

式(5.270)和(5.271)是反映薄膜 SOI MOSFET 中正、背栅之间电荷耦合作用的基本公式,联立二式可得到器件阈值电压与背栅偏压及器件参数之间关系的函数形式。

背面有积累、耗尽和反型三种状态,在反型状态时,即使 V_{Gf}低于阈值电压,器件仍然是导通的,实际电路中是不会采用这种方式的。下面讨论背面积累或耗尽时的情形。

背面积累状态 $\psi_{sb} \simeq 0$。在式(5.270)中令 $\psi_{sf} = 2\phi_F$ 和 $Q_{if} = 0$,可将阈值电压表示为

$$V_{T,ba} = V_{FBf} + (1 + C_s/C_{oxf})2\phi_F - Q_B/2C_{oxf} \tag{5.272}$$

相应地,背面达到积累状态时的栅压可令 $Q_{sb} = 0$,得

$$V_{Gb,a} = V_{FBb} - \frac{C_s}{C_{oxb}}2\phi_F - Q_B/2C_{oxb} \tag{5.273}$$

背面耗尽状态 $Q_{sb} = 0$,在式(5.270)中令 $\psi_{sb} = 2\phi_F$, $Q_{if} = 0$,即得

$$V_{T,bd} = V_{T,ba} - \frac{C_s C_{oxb}}{C_{oxf}(C_s + C_{oxb})}(V_{Gb} - V_{Gb,a}) \tag{5.274}$$

5.4.3 电流特性

假定硅膜的沟道区域均匀掺杂,采用缓变沟道近似,并忽略扩散电流分量。沿沟道方向流经 y 处的电流可表示为

$$I_D = -Z\mu_n Q_{if}(y)\mathrm{d}\psi_{sf}(y)/\mathrm{d}y \tag{5.275}$$

Z 是沟道宽度,μ_n 是沟道电子迁移率。从源($y = 0$)至漏($y = L$)积分上式,有

$$I_D = \frac{Z}{L}\mu_n \int_{2\phi_F}^{2\phi_F + V_{DS}} Q_{if}(y)\mathrm{d}\psi_{sf}(y) \tag{5.276}$$

由式(5.270)和(5.271)求出 $Q_{if}(y)$作为 $\psi_{sf}(y)$和 $\psi_{sb}(y)$或 ψ_{sf}和 $Q_{sb}(y)$的函数,代入上式并完成积分就可以算出漏极电流 I_D。

依赖于 V_{Gb},背面从源到漏会出现下述可能状态:全部耗尽,全部积累,部分积累部分耗尽,全部反型,部分反型部分耗尽,这里只讨论前二种背面状态下器件的电流特性。

背面全耗尽 $Q_{sb} = 0$,可以证明

$$I_{D,bd} = \frac{Z}{L}\mu_n C_{oxf}\left[(V_{Gf} - V_{T,bd})V_{DS} - (1 + \alpha_{bd})\frac{V_{DS}^2}{2}\right] \tag{5.277}$$

式中,

$$\alpha_{bd} = \frac{1}{C_{oxf}} \cdot \frac{C_s C_{oxb}}{C_s + C_{oxb}} \tag{5.278}$$

$C_s C_{oxb}/(C_s + C_{oxb})$是硅膜电容和背栅氧化层电容的串联值，$\alpha_{bd}$即为这个电容与正面栅氧化层电容之比。

对公式(5.277)取极值条件 $dI_D/dV_{DS}=0$，得到饱和漏极电流

$$I_{Dsat,bd} = \frac{Z\mu_n C_{oxf}}{2(1+\alpha_{bd})L}(V_{Gf} - V_{T,bd})^2 \tag{5.279}$$

相应的漏极饱和电压为

$$V_{Dsat,bd} = \frac{1}{1+\alpha_{bd}}(V_{Gf} - V_{T,bd}) \tag{5.280}$$

背面全积累

$$I_{D,ba} = \frac{Z}{L}\mu_n C_{oxf}\left[(V_{Gf} - V_{T,ba})V_{DS} - (1+\alpha_{ba})\frac{V_{DS}^2}{2}\right] \tag{5.281}$$

式中，

$$\alpha_{ba} = C_s/C_{oxf} \tag{5.282}$$

对公式(5.281)应用极值条件 $dI_D/dV_{DS}=0$，可得下述漏极电压时 I_D 开始饱和：

$$V_{Dsat,ba} = \frac{1}{1+\alpha_{ba}}(V_{Gf} - V_{T,ba}) \tag{5.283}$$

相应的饱和漏极电流为

$$I_{Dsat,ba} = \frac{Z\mu_n C_{oxf}}{2L(1+\alpha_{ba})}(V_{Gf} - V_{T,ba})^2 \tag{5.284}$$

可以看出，体硅(或厚膜)器件、背面全耗尽器件和背面全积累器件的漏极电流可用统一的公式表示如下：

$$I_{Dsat} = \frac{Z\mu_n C_{oxf}}{2L(1+\alpha)}(V_{Gf} - V_T)^2 \tag{5.285}$$

对体硅(或厚膜)器件，$\alpha_{si} = C_D/C_{oxf}$，$C_D = \varepsilon_s/X_d$，$X_d$ 是硅膜表面耗尽层最大厚度；对背面全耗尽薄膜器件，$\alpha_{bd} = \dfrac{C_s C_{oxb}}{C_{oxf}(Cs + C_{oxb})}$，$C_s = \varepsilon_s/t_s$，$t_s$ 是硅薄膜厚度，$t_s < X_d$；对背面全积累薄膜器件，$\alpha_{ba} = C_s/C_{oxf}$。显然，对上述三类器件，在栅氧化层厚度和硅膜掺杂浓度相同的情形下有

$$\alpha_{bd} < \alpha_{si} < \alpha_{ba} \tag{5.286}$$

就漏极饱和电流来说，从式(5.285)可知，在阈值电压和外加栅压相同的情形下，以背面全耗尽器件最大，体硅或厚膜器件次之，背面全积累薄膜器件最小。这样，背面全耗尽薄膜 SOI MOSFET 具有较大的电流驱动能力，相应的电路具有较好的速度特性。

5.4.4 亚阈区斜率(摆幅)

根据定义 $S = dV_G/d(\lg I_D)$及亚阈值电流 I_D 和栅压 V_G 的关系，对长沟道 MOSFET(体硅器件)，公式(5.96)表明，亚阈值斜率可统一表示为

$$S = (\ln 10)\phi_t dV_G/d\psi_s \tag{5.287}$$

式中 $\phi_t = k_B T/q$。对于 SOI MOSFET，可以设想有类似的 I_D 和 V_G 的关系，即这类器件的亚阈值斜率可以通过式(5.287)求得。

图 5.33 表示 MOS 器件的电容等效电路，由图 5.33(a)，不难看出对体硅(或厚膜)器件有

$$C_{ox}\mathrm{d}(V_G - \psi_s) = C_D\,\mathrm{d}\psi_s \qquad (5.288)$$

或

$$\mathrm{d}V_G/\mathrm{d}\psi_s = 1 + C_D/C_{ox} \qquad (5.289)$$

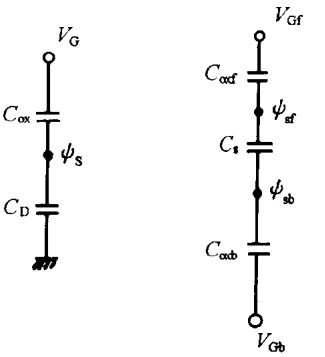
将式(5.289)代入式(5.287)，即得

$$S = \phi_t(\ln 10)(1 + C_D/C_{ox}) \qquad (5.290)$$

这一结果与式(5.95)一致。

对于背面全耗尽的 SOI 薄膜器件，由图 5.33(b)可以得到

$$\frac{\mathrm{d}V_{Gf}}{\mathrm{d}\psi_{sf}} = \frac{1}{C_{oxf}}\frac{C_sC_{oxb}}{C_s + C_{obx}} + 1 \qquad (5.291)$$



(a) 体硅或厚膜器件 (b) 背面全耗尽 SOI 薄膜器件

图 5.33 MOSFET 的电容分压电路

上式也可以通过正、背栅电荷耦合的基本公式(5.270)和(5.271)求出 V_{Gf}和 ψ_{sf}的函数关系以后得到。一般 $C_{oxb} \ll C_s$ 和 $C_{oxb} \ll C_{oxf}$，所以 $\mathrm{d}V_{Gf}/\mathrm{d}\psi_{sf} \simeq 1$，这时由式(5.287)有

$$S \simeq \phi_t \ln 10 \qquad (5.292)$$

将式(5.292)同式(5.290)比较可以看出，背面全耗尽薄膜 SOI 器件的亚阈值斜率(摆幅)比体硅或厚膜 SOI 器件要小。一般，如果忽略界面陷阱的影响，亚阈值斜率可统一表示为

$$S = \psi_t(\ln 10)(1 + \alpha) \qquad (5.293)$$

式中 α 的含义与前面所作的说明相同，并且服从下述关系

$$\alpha_{\text{背面全耗尽SOI器件}} < \alpha_{\text{体硅器件}} < \alpha_{\text{背面全积累SOI器件}}$$

因为背面全耗尽 SOI 薄膜 MOSFET 的亚阈值斜率(摆幅)最小，所以在用于实际电路时可以采用比较低的阈值电压而不会增加 $V_G=0$ 时的漏电流，这样可以得到较好的速度特性。

5.4.5 热电子效应

SOI MOSFET 存在 n^+-p-n^+结构，如果源、漏间距足够小，则寄生双极晶体管可以将基极电流(即漏端附近高场区碰撞电离所产生的空穴电流 I_h)放大，结果引起附加的电子电流 I_{bi}(即漏极电流增量)，如图 5.34 所示，其中 $I_h=(M-1)I_{ch}$，I_{ch}为进入高场区的电子电流，M 是电离倍增因子，$I_{bi}=\beta I_h=\beta(M-1)I_{ch}$。$\beta$ 是双极晶体管的电流增益。漏电流增加在器件中产生正反馈效应，如果高场区电场足够高，使乘积 $\beta(M-1)>1$，从而正反馈回路被触发，则器件无法关断。这一现象称为单晶体管的闩锁效应。

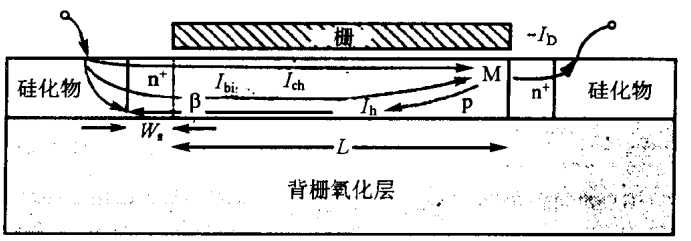


图 5.34 SOI 器件的寄生双极晶体管效应

SOI MOSFET 的击穿电压受基区浮空的寄生双极管支配，击穿条件为

$$\alpha_{npn}M = 1 \tag{5.294}$$

式中 α_{npn}为共基极电流增益,且

$$\alpha_{npn} \simeq 1 - W_B^2/2L_n^2 \tag{5.295}$$

W_B是基区宽度,可近似认为等于有效沟道长度 L,L_n 为电子扩散长度,倍增因子 M 可表示为

$$M = [1 - (BV_{CEO}/V_{Dsub})^n]^{-1} \tag{5.296}$$

BV_{CEO}是基区开路的共发射极击穿电压,V_{Dsub}是漏-衬底结的击穿电压,n 一般在 2～6 之间。从式(4.253)至(4.255)可得到 SOI MOSFET 的击穿电压为

$$V_B = BV_{CEO} \simeq \frac{V_{Dsub}}{(2)^{1/n}}\left(\frac{L}{L_n}\right)^{2/n} \tag{5.297}$$

上式表明,当 $L_n > L$ 时,即器件的沟道很短或 SOI 材料中电子寿命很长时,器件的击穿电压将降低。热电子的另一种效应是注入到栅氧化层中,使 SiO_2 界面受到破坏,器件的寿命 τ 可由下式表示:

$$\tau \propto \frac{Z}{I_D}(M-1)^{-m} \tag{5.298}$$

式中 $m \simeq 3$[可参见式(4.197)],碰撞电离倍增因子 M 可通过在漏端附近对电离率的积分得到,有

$$M - 1 = \frac{A_1}{B_1}[V_{DS} - V_{Dsat}]\exp\left(-\frac{B_1 L_c}{V_{DS} - V_{Dsat}}\right) \tag{5.299}$$

L_c 是特征长度,$L_c = t_s\left[\frac{\beta C_s}{2C_{oxf}(1+\alpha)}\right]$,背面全积累时 $\beta = 1$,背面全耗尽时 $\beta = 1 + \frac{C_s}{C_s + C_{oxf}}$。$A_1$ 和 B_1 是电子的碰撞电离率常数,可分别假设等于 $1.4 \times 10^6 cm^{-1}$和 $2.6 \times 10V/cm$。显然,可以预期全耗尽 SOI MOSFET 的热电子退化要弱一些。

5.5 埋沟 MOSFET

如果在 MOSFET 的沟道区注入和衬底导电类型相反的杂质,则可以构成一个埋沟器件。图 5.35(a)表示 n 沟道埋沟 MOSFET 的截面。由图可见,这种器件利用了体内迁移率高于表面迁移率的特点,导电沟道被移入半导体内,而不像常规 MOSFET 那样紧靠 SiO_2/Si 界面。埋沟器件一般做成耗尽型,但也可以做成增强型,取决于表面掺杂浓度和注入深度等。下面以 n 沟耗尽型为例分析这类器件的基本特性。

埋沟 MOSFET 和常规 MOSFET 不同之处在于沟道注入区和衬底形成 p-n 结,即沟道区内存在两个耗尽区,一个是表面处受栅压控制的表面耗尽区,另一个是受衬底偏压控制的注入区-衬底 p-n 结耗尽区,这两个耗尽区之间才是真正的导电沟道。

图 5.35(b)表示源端附近半导体内的情形,X_1 是表面耗尽区厚度,与表面势 ψ_s 的关系为

$$X_1 = \sqrt{\frac{2\varepsilon_s}{qN_i}(-\psi_s)} \tag{5.300}$$

其中 $N_i = N_I/X_I$,N_I 和 X_I 分别为注入剂量和注入深度,X_2 是注入区-衬底结的耗尽区在注入区中的厚度,由突变结理论可得:

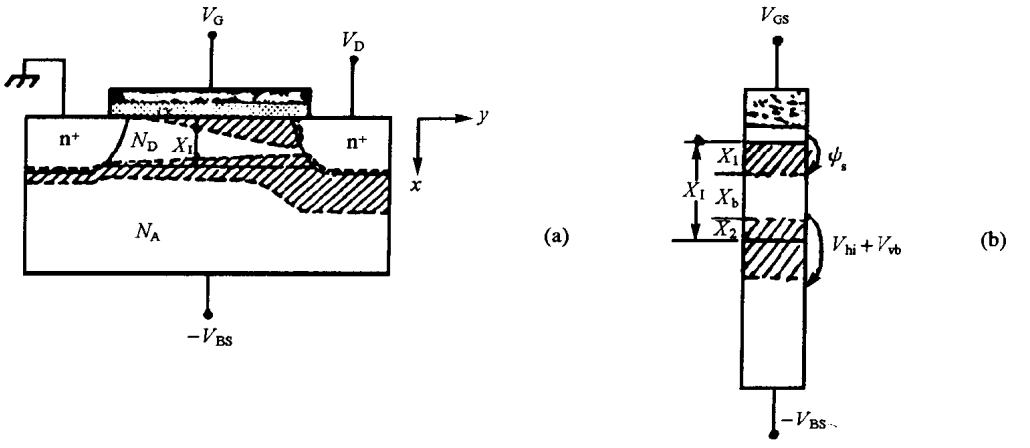


图 5.35　埋沟 MOSFET(a)及其源端附近区域(b)的截面图

$$X_2 = \sqrt{\frac{2\varepsilon_s}{q} \cdot \frac{N_A}{(N_i + N_A)N_i}(V_{bi} + V_{sB})} \tag{5.301}$$

V_{bi}是注入区-衬底结构的自建势，N_A是衬底(p 型)的掺杂浓度，在写 X_1 和 X_2 的表示式时对注入层杂质采取均匀分布近似，即假定 N_i 与注入深度无关。

$X_b = X_I - (X_1 + X_2)$是导电沟道厚度，当 $X_b = 0$ 时，沟道在源端被夹断，相应的栅压就是埋沟器件的阈值电压，常以 V_p 表示，与常规 MOSFET 阈值电压的组成结构类比，可得

$$V_p = V'_{FB} - \frac{qN_i}{2\varepsilon_s}(X_1 - X_2)^2 - \frac{qN_I}{C_{ox}}(X_1 - X_2) \tag{5.302}$$

上式中，$-\frac{qN_I}{2\varepsilon_s}(X_1 - X_2)^2$ 的项为夹断时的等效表面势，$-\frac{qN_I}{C_{ox}}(X_1 - X_2)$的项为夹断时表面耗尽层电荷在栅氧化层上产生的电压降(C_{ox}是单位面积栅氧化层电容)，F'_{FB}为栅、氧化层和注入层所成结构的平带电压，可通过没有注入层时的平带电压 V_{FB}表示为

$$V'_{FB} = V_{FB} + V_{bi} \tag{5.303}$$

利用夹断时的表面耗尽层电容

$$C_s = \frac{\varepsilon_s}{X_1} = \frac{\varepsilon_s}{X_I - X_2} \tag{5.304}$$

可将(5.302)改写为

$$V_p = V'_{FB} - \frac{qN_I}{C_{ox}}\left(1 + \frac{C_{ox}}{2C_s}\right) + \frac{1}{C_{ox}}\left(1 + \frac{C_{ox}}{2C_s}\right)\sqrt{\frac{2q\varepsilon_s N_i N_A}{(N_i + N_A)}(V_{bi} + V_{SB})} \tag{5.305}$$

导电沟道是电中性的，施主离子电荷被运动电子所中和，而电流则来自这些运动电子的贡献，以 Q_n 表示沟道区单位面积的电子电荷量，则

$$Q_n = -(qN_I - Q_s - Q_B) \tag{5.306}$$

式中，Q_B 是注入区-衬底结边耗尽层单位面积的电荷量，数值(设 $N_i \gg N_A$)为

$$Q_B = \sqrt{2q\varepsilon_s N_A(V_{bi} + V_{SB} + V(y))} \tag{5.307}$$

Q_s 为表面耗尽层单位面积电荷量，近似为

$$Q_s = C_{ox}[V_{GS} - V'_{FB} + V(y)] \tag{5.308}$$

漏极电流可由下式求得

$$I_D = \frac{Z}{L}\int_0^{V_{DS}} \mu_B(-Q_n)\mathrm{d}V \tag{5.309}$$

μ_B 为体内迁移率，I_D 的最终表达式为

$$I_D = \frac{Z}{L}\mu_B\left\{qN_I V_{DS} + C_{ox}\left[(V_{GS} - V'_{FB})V_{DS} - \frac{1}{2}V_{DS}^2\right]\right.$$
$$\left. - \frac{2}{3}\sqrt{2q\epsilon_s N_A}\left[(V_{bi} + V_{DS} + V_{SB})^{\frac{3}{2}} - (V_{bi} + V_{SB})^{\frac{3}{2}}\right]\right\} \tag{5.310}$$

5.6 电荷耦合器件(CCD)

CCD 的基本结构是一种密排的 MOS 二极管阵列。工作时，信息量用电荷(简称电荷包)代表，而不是常规器件惯用的电流量或电压量。MOS 二极管将被偏置到深耗尽状态，以实现信息电荷的储存。如果脉冲是按一定位相次序加到 MOS 二极管阵列上的时钟脉冲，信息电荷将沿衬底表面转移。CCD 被用于摄像、信息处理和数字存储，用于摄像系统的 CCD 采用光注入，其余二者采用电注入。CCD 分为表面沟道 CCD(SCCD)和埋沟 CCD(BCCD)，在 SCDD 中，电荷在半导体表面存储和转移，在 BCCD 中，半导体表面掺有和衬底相反类型的杂质，因此电荷存储和转移发生在紧靠表面的半导体内部。这里以 SCCD 为例，讨论 CCD 的基本原理。

5.6.1 表面深耗尽状态和电荷存储

深耗尽指加在二极管上的电压(栅压)V_G 超过发生强反型的阈值 V_T 时并没有反型层形成的半导体表面状态。以下考虑 p 型硅衬底上的 MOS 二极管。由于反型层(电子层)夹在其上的栅氧化层和其下的耗尽层之间，电子只能来自耗尽层内的热产生过程，所以，当 $V_G > V_T$ 时，必须经历一个时间为 τ_T 的弛豫过程才能在表面形成反型层，而在小于 τ_T 的时间范围内，半导体表面仍处在耗尽状态，即深耗尽状态。深耗尽是一种非平衡状态，可以通过耗尽层内热产生的最大速率 G_{max}来估计 τ_T。以 X_d 和 X_{dm}分别表示耗尽时和达到强反型(热平衡状态)时的耗尽层厚度，则有

$$G_{max}\tau_T X_d = N_A(X_d - X_{dm}) \simeq N_A X_d \tag{5.311}$$

所以，

$$\tau_T \simeq \frac{N_A}{G_{max}} = \frac{2\tau_0 N_A}{n_i} \tag{5.312}$$

N_A 是衬底的掺杂浓度，τ_0 是电子的有效寿命。如果 τ_0 是微秒的数量级，τ_T 就将是秒的量级。τ_T 越大，表示半导体表面状态由深耗尽到强反型所需要的时间越长。对 CCD，需要半导体表面处在深耗尽状态，所以 τ_T 越大越好。

图 5.36 表示 p 型衬底 MOS 二极管在半导体表面深耗尽时的能带图；也表示了半导体表面电势极小值所形成的电子(少子)势阱深度。其中，图 5.36(a)对应信息电荷 $Q_{sig}=0$ 的情形，这时势阱是空着的，图 5.36(b)对应势阱中存储了信息电荷($Q_{sig}\neq 0$)的情形，这时表面势减少，势阱深度变化。显然，存储的最大信息电荷所引起势阱深度变化不得超过原来的势阱深度 $q\psi_{s0}$，否则会

溢出势阱，造成信息失真。在图中，V_G 是相对衬底的栅压，ψ_s 是表面势，即势阱深度。

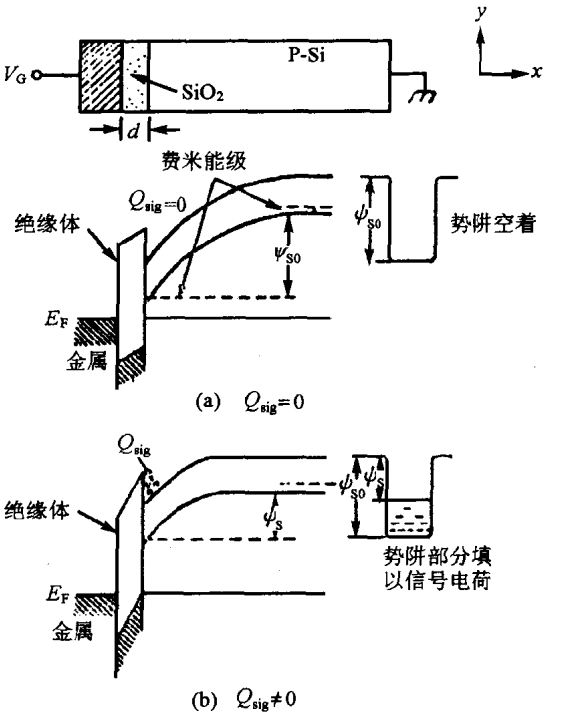


图 5.36　表面沟道 MOS 二极管深耗尽时的能带图及势阱深度

为了得到势阱深度和栅压及信息电荷量之间的关系，将公式(5.30)重写如下：

$$V_G = V_{FB} + \psi_s - Q_s/C_{ox} \tag{5.313}$$

其中 $Q_s = Q_B + Q_{sig}$，Q_{sig} 是信息电荷，Q_B 是耗尽层电荷，它们都相对单位面积半导体表面而言。Q_B 可表示为

$$Q_B = -qN_AX_d = -(2q\epsilon_sN_A\psi_s)^{\frac{1}{2}} \tag{5.314}$$

以 ψ_{so} 表示 $Q_{sig}=0$ 时的表面势，则根据式(5.313)和(5.314)，有

$$\psi_{so} = V_G - V_{FB} - B[\sqrt{1+2(V_G-V_{FB})/B} - 1] \tag{5.315}$$

$$B = q\epsilon_sN_A/C_{ox}^2 \tag{5.316}$$

式(5.315)是设计 CCD 的一个重要公式。

以 ψ_s 表示势阱信息电荷 Q_{sig} 部分填充时的表面势，则在式(4.297)中以 $V = V_G - V_{FB} - |Q_{sig}|/C_{ox}$ 代替 $V_G - V_{FB}$，即得

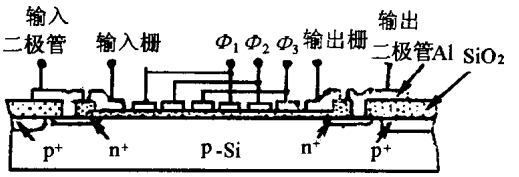
$$\psi_S = V - B[\sqrt{1+2V/B} - 1] \tag{5.317}$$

V 中的信息电荷 Q_{sig} 加了绝对值符号是因为 Q_{sig} 代表电子电荷，它本身之值为负，由式(5.317)可知当势阱中存储信息电荷时，表面势减小，即势阱深度减小。

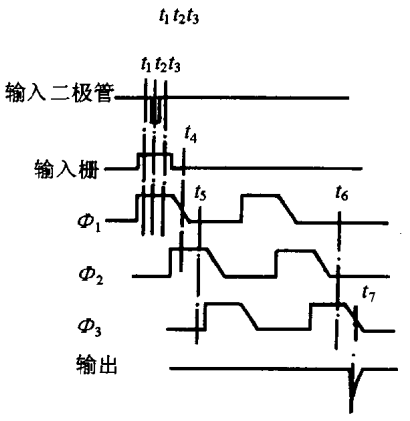
5.6.2　基本的 CCD 结构

图 5.37(a)表示出了三相 n 沟道 CCD 及其基本的输入、输出结构。接到 Φ_1，Φ_2 和 Φ_3 时钟

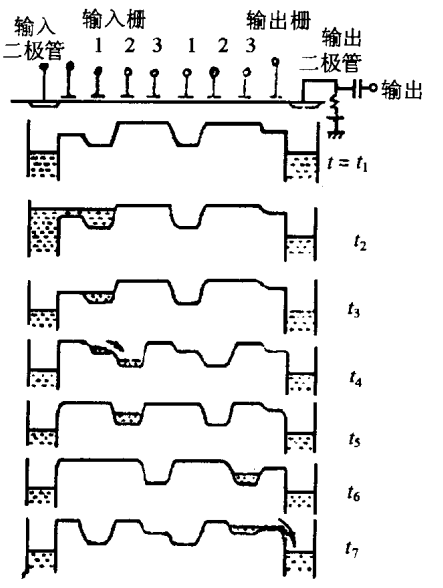
线的 6 个 MOS 二极管或电极构成了 CCD 的主体；输入二极管、输入栅、输出二极管和输出栅是向 CCD 主体注入电荷包以及从 CCD 主体检测电荷包的元件。等效地讲，CCD 的主体是在 MOSFET 的源极和漏极之间的密排的 MOS 电容阵列。在一般的大规模集成电路中，是通过布线将多数的 MOSFET 连接起来传递电信号的；而 CCD 的特点是使各个 MOS 电容的栅极电压三相化，不经过布线，由对应于源极的输入部向对应于漏极的输出部输送电荷。图5.37(b)示出了 CCD 的时钟波形，图 5.37(c)示出了相应的势阱和电荷分布。在 t_1 时刻，时钟线 Φ_1 处于高电压下，Φ_2 和 Φ_3 处于低电压下，因此 Φ_1 下的势阱比 Φ_2 和 Φ_3 的势阱要深。但此时输入二极管和输出二极管都被偏置于很高的正电压下以防止输入栅和输出栅下面的表面反型，并且输入二极管和输出二极管不能向 CCD 主阵列提供电子，即 CCD 阵列下面的势阱全部是空的。在时刻 t_2，输入二极管的电压降低，使得电子能经过输入栅注入(流)到第一个 Φ_1 电极下的势阱。在注入结束时，输入栅和 Φ_1 电极下的表面势将与输入二极管电势相同，此时，电子存储在输入栅和第一个 Φ_1 电极下。在 $t = t_3$，输入二极管的电压返回到高值，输入栅下的电子和第一个 Φ_1 电极下的过剩电子将经过输入二极管的引线流出器件，建立起在第一个 Φ_1 电极下轮廓分明的电荷包。在 $t = t_4$，加在 Φ_1 上的电压处于低值，而 Φ_2 上则加上了高电压，由于 Φ_2 下的势阱较深，存储在 Φ_1 下的电子向 Φ_2 电极下转移，由于载流子输运一段电极宽度的距离要有有限的时间，故加在 Φ_1 上的电压逐渐下降。在 $t = t_5$，电荷转移过程完成，原先的电荷包现在存储在第一个 Φ_2 电极下。重复此过程，在 $t = t_6$，注入的电荷包存储在第二个 Φ_3 电极下。在 $t = t_7$，Φ_3 电极的电压返回到低值，将电子推向输出二极管，从而在输出端得到正比于电荷包大小的输出信号。



(a) 基本结构



(b) 时钟波形和输入、输出信息



(c) 电势能和电荷分布

图 5.37　三相 n 沟 CCD

对于信息处理器件和存储器件，电荷包是用上述方法将适当的电压加到 CCD 输入端的 p-n 结而引入的。对于摄像系统应用，电荷包可以是输入二极管作为光探测器在入射光作用下产生的，也可以是入射光图像直接照射半导体衬底产生电子-空穴对而其中的少子(电子)被电极吸引而聚集于势阱中形成的。但是，后者因为光要穿过金属电极，所以入射光的损耗大，这是一个很大的缺点。因此，现在的光探测部分都采用光电二极管作为像素单位，使入射光图像成像在由许多微小光电二极管排列的平面阵列上。

电极结构对转移效率有一定影响。如果邻近电极之间存在势垒，即使时钟频率很低，信息电荷也不能完全转移出去。因此，必须从器件电极结构的设计考虑，尽量减小甚至消除极间势垒。现在一般采用交叠栅结构，使极间的距离很小。图 5.38 和 5.39 表示常用的两种交迭栅结构。

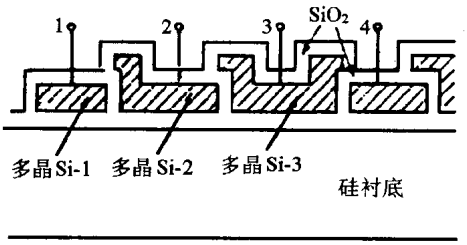


图 5.38　三重多晶硅栅 CCD，图中数字表示多晶硅淀积层次

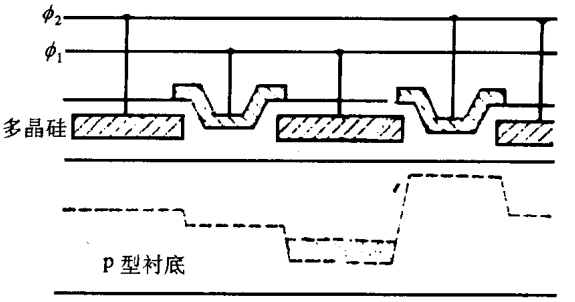


图 5.39　多晶硅-铝栅结构 CCD

5.6.3　转移效率和频率响应

对 CCD，一个最重要的参数是转移效率。在电荷由一个电极向另一个电极转移的过程中，由于种种原因，不可能把全部信息电荷转移出去，总会留下一小部分电荷。转移到下一个势阱中的电荷与原有的信息电荷的比值称为转移效率 η，残留电荷所占的比率为失真率 ε，显然

$$\eta + \varepsilon = 1 \tag{5.318}$$

当信息电荷转移过 N 个电极之后，总的转移效率为 η^N。对于 ε 很小的情况下，有

$$\eta^N = (1-\varepsilon)^N \simeq e^{-\varepsilon N} \tag{5.319}$$

对于实际的 CCD，总的转移次数往往大于 1000(即 $N>1000$)。为了保证经过 N 次转移以后总的转移效率 $\eta^N>90\%$，失真率应为 $\varepsilon<10^{-4}$。

电荷在电极之间的转移，不外是通过漂移和扩散二种运动形式来实现。引起电子漂移运动的电场包括二部分：一部分是势阱中信息电荷分布不均匀所引起的自感应电场，一部分是邻近栅极电压所引起的边缘电场。如前所述，信息电荷使表面势减小，信息电荷在势阱中分布的梯度将引起表面势有一个相反的分布梯度，即自感应电场与势阱中电子浓度变化的方向相反，它把电子推向下一个电极。仔细分析表明，在信息电荷转移的开始阶段，自感应电场所引起的漂移运动起主要作用。随着电场的转移，电荷数量不断减小。自感应电场的作用也越来越小，尽管绝大部分信息电荷的转移靠自感应电场的作用，但它并不决定最终电荷转移的弛豫

时间,决定弛豫时间的是剩余少量电荷的扩散运动和边缘电场所引起的漂移运动。热扩散运动所决定的弛豫时间为

$$\tau = \frac{4L^2}{\pi^2 D_n} \simeq \frac{L^2}{2.5D_n} \tag{5.320}$$

式中 L 为电极长度,D_n 为电子的扩散系数。也就是说,如果只考虑扩散运动,原来储存在势阱中的电荷随时间指数衰减:

$$\frac{Q(t)}{Q_0} = e^{-t/\tau} \tag{5.321}$$

$Q(t)$是 t 时刻残留在势阱中的电荷。如果需求失真率 $\varepsilon < 10^{-4} \simeq e^{-10}$,则要求时钟变化周期 T 满足

$$e^{-t/\tau} \leqslant e^{-10} \tag{5.322}$$

所以

$$T \geqslant 10\tau \tag{5.323}$$

举一个例子,若 $L = 10\mu m$, $D_n = 100 cm^2/s$, 则 $\tau = 4 \times 10^{-7} s$, 故三相时钟脉冲的最高频率是 $f = 2.5MHz$。也就是说,选择的时钟频率不能高于 2.5MHz。当然,时钟频率也不是越低越好。因为 CCD 是利用半导体表面深耗尽的器件,如果时钟频率太低,热产生电荷以及从耗尽层边缘扩散过来的电荷等会使失真率增大。

边缘电场可以加速电荷的转移,这是因为邻近电极的栅压对信息电荷具有吸引作用。所谓边缘场,是因为栅极电荷发出的电力线并不都局限于栅极的面积范围内,而是有一部分超出栅极的边缘。显然,氧化层越厚和半导体表面耗尽层越厚(衬底掺杂浓度越低)时,超出电极边缘的扩展越显著,边缘场作用越大。当然,增加栅压,减小电极间距和电极长度都可能增大边缘场的作用。对于典型情况,衬底掺杂浓度为 $10^{15}/cm^3$,栅氧化层厚度为 1000Å,栅压为 10 伏,由于边缘场加速电荷转移的作用,失真率下降,时钟频率可以提高到 10MHz。

此外,实际器件总存在着 Si/SiO_2 界面陷阱,它们可能俘获电荷,在信息电荷的转移过程中,这些被俘获的电荷中,有些可能被及时释放出来,跟上信息电荷的转移;有些则落在后面,造成信息电荷的损失。克服界面陷阱影响的方式是采用"胖零"工作模式,就是不管有无信息电荷,都让半导体表面储存一定的本底电荷,使界面陷阱基本被填满。

第六章　有源微波二极管

微波频率覆盖范围从 1GHz(10^9Hz)至 1000GHz,相应的波长从 30～0.03cm。其中 30 至 300GHz 频段,因其波长从 10～1mm,故称毫米波段;更高的频率称为亚毫米波段。

在前面讨论过的许多半导体器件,例如双极型晶体管、MESFET,HEMT 和 MOSFET 等,都可以产生微波信号或把微波信号放大,即都可以成为有源微波器件,但必须缩小器件尺寸并将寄生电容和电阻减到最小程度,以获得微波性能。显然,有源微波二极管具有结构简单的优点。本章考虑四类重要的有源微波二极管:隧道二极管、共振隧穿二极管、IMPATT 二极管、转移电子器件。我们主要讨论这些器件的工作原理。

6.1　隧道二极管

隧道二极管是最早出现的半导体量子器件,1958 年由江崎(Esaki)发明。这是一种利用量子隧道穿透现象的器件,电流来自多数载流子的运动。不过,载流子通过势垒的隧穿时间不受经典的渡越时间概念(势垒宽度除以载流子速度)支配,而受量子跃迁速率支配,反比于单位时间的量子跃迁几率。这一隧穿时间很短,使器件能在毫米波段工作。隧道二极管制造工艺简单,噪声低,电流峰谷比大,适合用在小功率微波振荡或放大电路中。

6.1.1　隧道输运过程

作为理解包括隧道二极管在内的隧道器件工作原理的基础,这一小节讨论半导体器件中的隧道输运过程。实际的半导体器件是由多层不同材料紧密接触而成的,包括:不同半导体材料的接触(S-S 结,例如 p-n 结),金属和半导体的接触(M-S 结),以及两种材料之间夹有绝缘层的结构(SIS 结、MIS 结、MIM 结)。这些结形成的势垒阻碍载流子通过,通常只有能量足够高的载流子才能越过。但是,如果材料掺杂浓度很高或外加电场很强以致势垒宽度很小,或者如果绝缘层(I 层)厚度降到几十个埃以下,则载流子有显著的几率直接穿过势垒,这就是隧穿输运过程。隧穿过程可以是直接过程,也可以是间接过程。直接隧穿过程中电子的动量 $\hbar\boldsymbol{k}$ 保持不变,即没有声子参与;间接隧穿则必须有声子参与,才能满足动量守恒和能量守恒条件。一般说来,间接隧穿的几率比直接隧穿的几率要低得多。我们的讨论限于直接隧穿过程。

1. 隧穿电流密度

设电子贯穿势垒的方向为 x 方向,则从势垒一侧入射到势垒上的电流密度为

$$J = \frac{2}{(2\pi)^3}\int(-q)v_x f_1(E(\boldsymbol{k}))\mathrm{d}\boldsymbol{k} \tag{6.1}$$

式中 $\mathrm{d}\boldsymbol{k} = \mathrm{d}k_x\mathrm{d}k_y\mathrm{d}k_z$ 为 $\boldsymbol{k}$ 空间的体积元,$f_1(E(\boldsymbol{k}))$为入射一侧电子的费米分布函数,v_x 为电子在 x 方向的运动速度,q 为电子的电荷量,因子$(2\pi)^3$ 和 2 分别表示对于单位实际空间体积($V=1$)每一个允许状态(量子态)在 $\boldsymbol{k}$ 空间所占体积和电子有两种自旋状态。

我们采用球形等能面近似。这样,晶体中的电子与自由电子不同之处只是应当用有效质

量 m。根据德布罗意关系,电子动量为

$$\boldsymbol{P} = \hbar \boldsymbol{k} \tag{6.2}$$

$\hbar = h/2\pi$, h 是普朗克常数。电子的能量 $E(\boldsymbol{k})$和速度 $v(\boldsymbol{k})$分别为

$$E(\boldsymbol{k}) = \frac{\hbar^2 k^2}{2m} \tag{6.3}$$

$$v(\boldsymbol{k}) = \frac{\hbar \boldsymbol{k}}{m} \tag{6.4}$$

利用以上 3 式,可以将式(6.1)改写为

$$J = -\frac{4\pi q m}{h^3}\int f_1(E(\boldsymbol{k}))\mathrm{d}E_x\mathrm{d}E_\perp \tag{6.5}$$

$E_\perp$是与垂直于隧穿方向的动量(横向动量)相关的能量,电子在垂直于隧穿方向的平面内运动不受约束;E_x 是与隧穿方向的动量相关的能量。E_x 和 $E_\perp$分别为

$$E_x = \frac{\hbar^2 k_x^2}{2m} \quad , \quad E_\perp = \frac{\hbar^2(k_y^2 + k_z^2)}{2m} \tag{6.6}$$

$$E = E_x + E_\perp \tag{6.7}$$

入射电子中只有少数能够通过势垒。以 T_t 表示势垒另一侧的状态全空时的隧穿几率,则该侧的状态被部分占据时的隧穿几率为 $T_t[1 - f_2(E(\boldsymbol{k}))]$, $f_2(E(\boldsymbol{k}))$是势垒另一侧的费米分布函数。这样,电子从势垒一侧到另一侧的隧穿电流密度为

$$J_{1\to 2} = -\frac{4\pi q m}{h^3}\int f_1(E)T_t[1 - f_2(E)]\mathrm{d}E_x\mathrm{d}E_\perp \tag{6.8}$$

同样,如果假定隧穿几率 T_t 在相反的两个方向上相同,则相反方向的隧穿电流为

$$J_{2\to 1} = -\frac{4\pi q m}{h^3}\int f_2(E)T_t[1 - f_1(E)]\mathrm{d}E_x\mathrm{d}E_\perp \tag{6.9}$$

当势垒加偏压时,由隧穿过程引起的电流密度为

$$\begin{aligned} J_t &= J_{1\to 2} - J_{2\to 1} \\ &= \frac{4\pi q m}{h^3}\int T_t[f_2(E) - f_1(E)]\mathrm{d}E_x\mathrm{d}E_\perp \end{aligned} \tag{6.10}$$

$f_1(E)$, $f_2(E)$是势垒两侧的费米分布函数。此外,由式(6.7)可知,E, E_x 和 $E_\perp$三者可任选二个作为独立变量,但要注意相应的积分范围。

2. 隧穿几率

根据量子力学中的准经典近似(*W.K.B* 方法),隧穿几率可表示为

$$T_t = \mathrm{e}^{-2\int_{-x_1}^{x_2} |k| \mathrm{d}x} \tag{6.11}$$

$$|k| = \sqrt{\frac{2m}{\hbar^2}[U(x) - E]} = \sqrt{\frac{2m}{\hbar^2}E_b(x)} \tag{6.12}$$

$|k|$是势垒中载流子波矢的绝对值,其中 $U(x)$为势能,E 为载流子进入势垒的能量,在经典转折点 $-x_1$ 和 x_2 处 $U \simeq E$, $E_b(x)$称为有效势垒。

在半导体器件中,隧穿电子所遇到的势垒有三种基本模型:矩形势垒、三角形势垒和抛物形势垒。矩形势垒的有效高度是常数,情形比较简单,不在此讨论。下面以 p-n 结中的隧穿过程为例,分析三角形势垒和抛物形势垒的隧穿几率。图 6.1 表示 p-n 结的直接隧穿现象及有

效势垒模型。

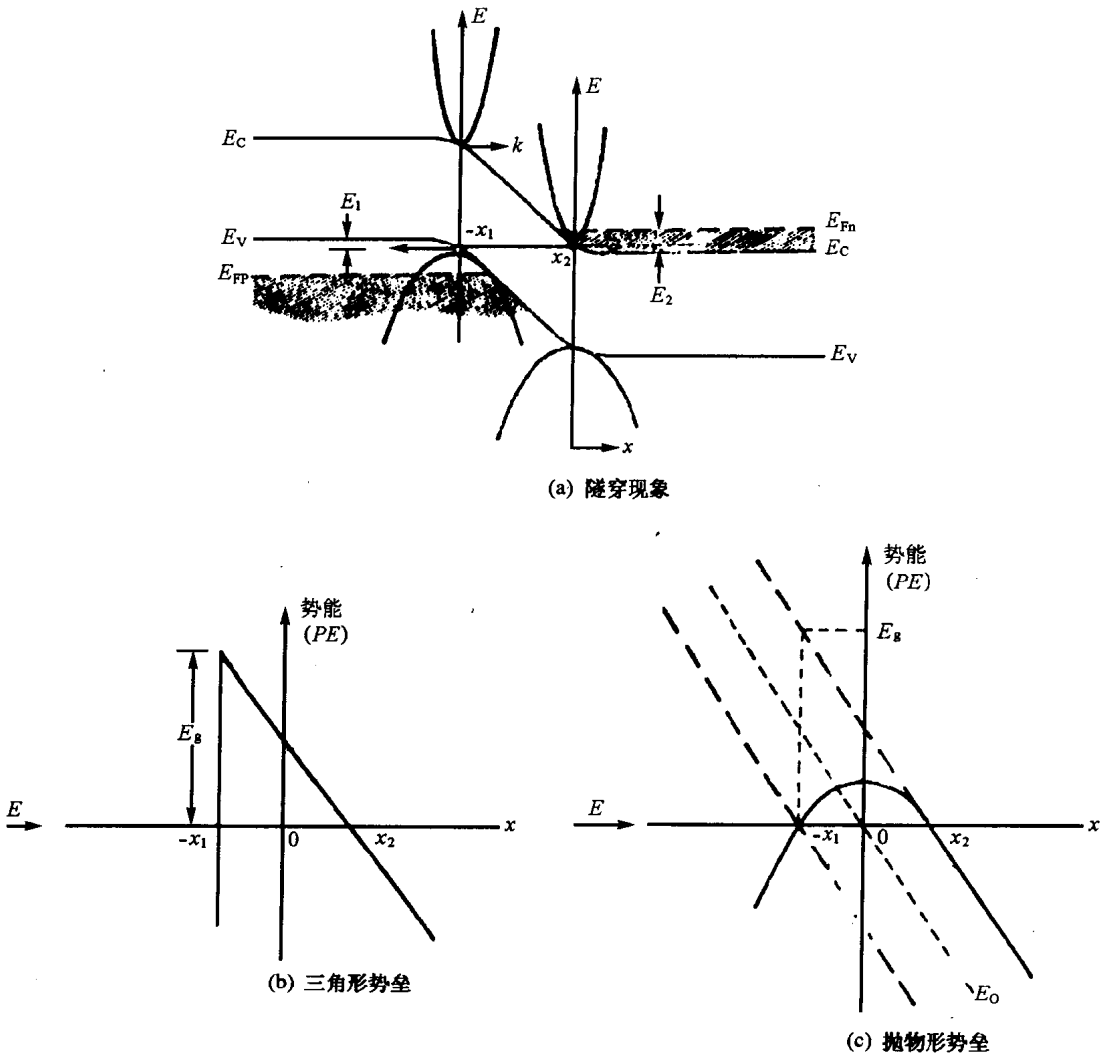


图 6.1　p-n 结中的隧穿现象和有效势垒模型

对三角形势垒,有效势垒为

$$E_b(x) = \frac{E_g}{2} - q\mathscr{E}x \tag{6.13}$$

式中,E_g 为半导体的禁带宽度,$\mathscr{E}$ 为电场,在 $x = x_2$ 处,

$$\frac{E_g}{2} - q\mathscr{E}x = 0$$

以及在 $x = -x_1$ 处,

$$\frac{E_g}{2} - q\mathscr{E}x = E_g$$

所以由式(6.11)可得

$$T_t = e^{-\frac{4\sqrt{2}}{3}\cdot\frac{m^{1/2}E_g^{3/2}}{q\hbar\mathscr{E}}} \tag{6.14}$$

对于抛物形势垒，在经典转折点($-x_1$ 和 x_2)处势垒连续，不像三角形势垒那样在 $-x_1$ 处发生跳变。这种势垒模型的有效势垒为

$$E_b(x) = \frac{(E_g/2)^2 - (q\mathscr{E}x)^2}{E_g} \tag{6.15}$$

由 $(E_g/2)^2-(q\mathscr{E}x)^2=0$，得到经典转折点的位置为

$$-x_1, x_2 = \mp\frac{E_g}{2q\mathscr{E}}$$

所以由式(6.11)，得到隧穿几率为

$$T_t = e^{-\frac{\pi}{2\sqrt{2}}\cdot\frac{m^{1/2}E_g^{3/2}}{q\hbar\mathscr{E}}} \tag{6.16}$$

除数值常数外，上式和表示三角形势垒隧穿几率的公式(6.14)实际上相同。

因为在隧穿过程中总动量必须守恒，所以在三维情形下，虽然隧穿仍沿 x 方向，但必须考虑横向动量的影响。以抛物形势垒为例，这时有

$$\frac{\hbar^2k_x^2}{2m} + E_\perp = -E_b(x) = -\frac{(E_g/2)^2-(q\mathscr{E}x)^2}{E_g} \tag{6.17}$$

由上式得

$$|k_x| = \sqrt{\frac{2m}{\hbar^2}\cdot\frac{(E_g/2)^2 + E_\perp E_g - (q\mathscr{E}x)^2}{E_g}} \tag{6.18}$$

由于隧穿沿 x 方向，隧穿几率应表示为

$$T_t = e^{-2\int_{-x'_1}^{x'_2}|k_x|\,dx} \tag{6.19}$$

$-x'_1$ 和 x'_2 由方程 $(E_g/2)^2+E_\perp E_g-(q\mathscr{E}x)^2=0$ 决定，为

$$-x'_1, x'_2 = \mp\frac{1}{2q\mathscr{E}}\sqrt{E_g^2+4E_\perp E_g} \tag{6.20}$$

将式(6.18)和(6.20)代入式(6.19)，完成积分，得

$$\begin{aligned} T_t &= e^{-\frac{\pi}{2\sqrt{2}}\cdot\frac{m^{1/2}E_g^{3/2}}{q\hbar\mathscr{E}}}\cdot e^{-\pi(2mE_g)^{1/2}E_\perp/q\hbar\mathscr{E}} \\ &= e^{-E_g/4\overline{E}}\cdot e^{-E_\perp/\overline{E}} \end{aligned} \tag{6.21}$$

式中，第一个指数对应横向动量为零的隧穿几率；第二个指数表征横向动量的影响，它使隧穿几率减小。$\overline{E}$ 为

$$\overline{E} = \frac{q\hbar\mathscr{E}}{\pi(2mE_g)^{1/2}} \tag{6.22}$$

若 $\overline{E}$ 很小，只有横向动量小的电子能够隧穿。从上述结果可见，要得到大的隧穿几率，有效质量 m 和禁带隙 E_g 应很小，电场 $\mathscr{E}$ 应很大。

3. *反向隧穿电流与隧道击穿*

在反偏情形下，p 区的能带相对 n 区升高，价带电子可穿过禁带，到达 n 区的导带内，相应有电流流过 p-n 结。图 6.2 表示这种反向隧道过程。

根据图 6.2 中费米能级相对能带边的位置，可以认为 p 区价带全满和 n 区导带全空，即

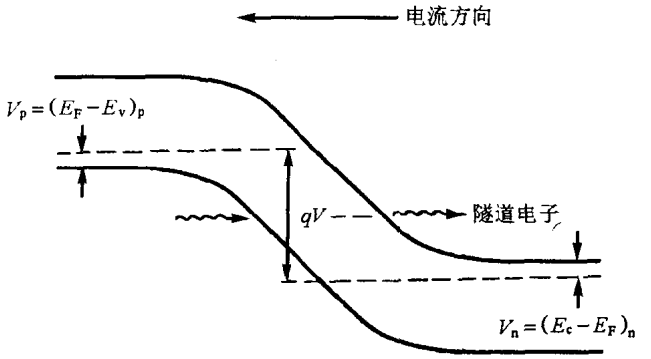


图 6.2　反向偏置 p-n 结的能带图

$f_1(E)=f_v(E)=1$ 和 $f_2(E)=f_c(E)=0$，于是计算隧穿电流的公式(6.10)简化为

$$J_t=\frac{4\pi qm}{h^3}\iint T_t\,dE_x dE_\perp \tag{6.23}$$

如果采用抛物形势垒近似并考虑横向动量影响，则隧穿几率 T_t 由公式(6.21)给出；电子在垂直于隧穿方向的平面内运动不受约束，所以 $E_\perp$ 的取值范围是$(0,\infty)$，在将式(6.21)代入式(6.23)并完成对 $E_\perp$ 的积分以后得到

$$J_t=\frac{4\pi qm}{h^3}\overline{E}e^{-E_g/4\overline{E}}\int dE_x \tag{6.24}$$

E_x 的取值范围由能量守恒条件决定。由图 6.2 不难看出，只有能量介于 p 区价带顶 E_{vp} 和 n 区导带底 E_{cn} 之间的电子能够参与隧穿过程，即 E_x 的取值范围是 E_{cn} 至 E_{vp}，这一范围的大小为 $qV-qV_p-qV_n$，V 是外加反向电压，qV_p 和 qV_n 分别为 P 区和 n 区内费米能级和能带边之间的能量差($qV_p=E_{Fp}-E_{vp}$，$qV_n=E_{cn}-E_{Fn}$)。对简并 p-n 结，费米能级 E_{Fp} 和 E_{Fn} 进入各自的能带并接近能带边，故 qV_p，qV_n 与 V 相比很小而可以略去，这时 E_x 的取值范围近似为 qV。这样，我们最终得到通过 p-n 结的反向隧穿电流密度近似为

$$J_t=\frac{4\pi qm}{h^3}(qV)\overline{E}\,e^{-E_g/4\overline{E}} \tag{6.25}$$

在公式(6.25)中，J_t 和电场强度 $\mathscr{E}$ 的关系反映在参量 $\overline{E}$ 中[见式(6.22)]。为了明确，可以将其改写为

$$J_t=\frac{\pi q^2 m}{h^3}VE_g\frac{\mathscr{E}}{\mathscr{E}_0}e^{-\mathscr{E}_0/\mathscr{E}} \tag{6.26}$$

$$\mathscr{E}_0=\frac{\pi(2m)^{1/2}E_g^{3/2}}{4q\hbar} \tag{6.27}$$

显然，对一定的半导体材料 $\mathscr{E}_0$ 为常数。例如对 GaAs，$E_g=1.42$eV 和 $m=0.068m_0$(m_0 是自由电子质量)，$\mathscr{E}_0\simeq 1.7\times10^7$V/cm。

由(6.26)可知，p-n 结的反向隧穿电流强烈依赖于电场强度 $\mathscr{E}$，随 $\mathscr{E}$ 增大而指数增加。当反向偏压 V(或电场 $\mathscr{E}$)增大到一定程度，隧道电流将大大超过热扩散引起的反向饱和电流，这就是所谓隧道击穿。对于高掺杂的 p-n 结，由于电子沿隧穿方向运动的能量 E_x 可以在较大范围内取值，容易发生隧道击穿。此外，在得到式(6.26)的过程中，假定了电场强度 $\mathscr{E}$(通过 $\overline{E}$)

为常数，实际计算时可以取平均值$\frac{1}{2}\mathscr{E}_m$来代替$\mathscr{E}$；$\mathscr{E}_m$是 p-n 结内的最大电场强度，对突变 p-n 结有

$$\mathscr{E}_m = \sqrt{\frac{2qN}{\varepsilon_s}(V + V_{bi})} \tag{6.28}$$

其中 $N^{-1} = N_A^{-1} + N_D^{-1}$，$N_A$ 和 N_D 分别为 p 区和 n 区的掺杂浓度。

6.1.2 隧道二极管

隧道二极管由 p 区和 n 区都是高掺杂(掺杂浓度 $10^{19}cm^{-3}$左右)的简并 p-n 结组成，图6.3是它在热平衡时的能带图。简并量 qV_p 和 qV_n 的典型值为几个 k_BT，耗尽层势垒的宽度为100Å左右。加上电压以后，电子有一定的隧穿几率穿过禁带，从导带到价带，或从价带到导带。

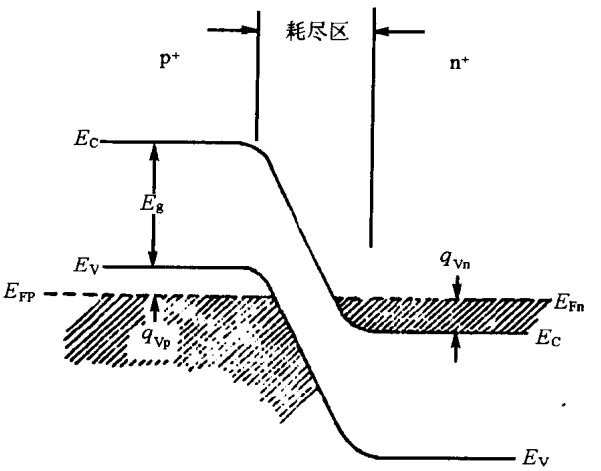


图 6.3 隧道二极管在热平衡状态下的能带图

1. 电流-电压特性

图 6.4 表示隧道二极管典型的静态电流-电压特性，我们可用该图中的插图对它解释如下。当施加正向电压时，存在一个能态“交叠区”，它在 n 型一侧为填满电子的能态，与 p 一侧的可用空态相对应，因而电子可以从 n 型一侧隧穿到 p 一侧。交叠区的能量范围起初随外加电压 V 增加，在 $V = V_p$ 时达到最大；$V > V_n$ 时开始缩小；直至 $V = V_p + V_n$ 时下降到零，这时 n 型一侧的导带底正对着 p 型一侧的价带顶，没有可用的空态与填满的能态相对，而不再流过隧道电流。因此可以预期，随着正向电压 V 的增加，隧道电流从零增加到峰值 I_p，然后在 $V = V_p + V_n$ 时减少到零；V_p 和 V_n 分别为 p 区和 n 区的简并量(参见图 6.3)。电压更高时，将流过正常的由少数载流子注入引起的热扩散电流，它随 V 指数增加。在实际的隧道二极管中，由于耗尽层内存在杂质或缺陷引起的深能级，电子可以经由这些能态跃迁到价带，这种“间接”隧穿过程导致称之为“过剩电流”的附加电流，因此正向电流的极小值(谷值)并不为零。隧道二极管电流-电压特性的一个重要特点是从峰值电流 I_p 降到谷值电流 I_V 的阶段形成了负阻区，故这种器件可用作电子线路中的放大器或振荡器，其效率和比值 I_p/I_V 有关。

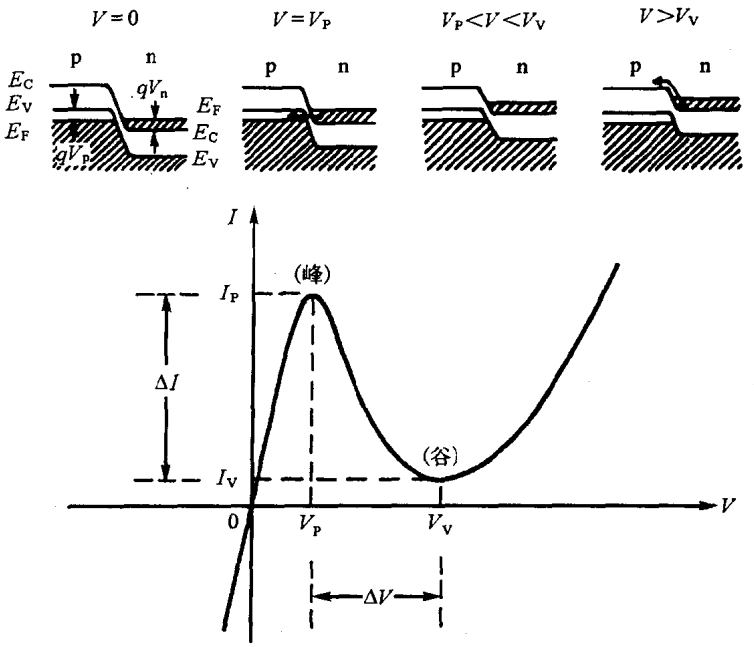


图 6.4　隧道二极管典型的静态电流-电压特性，插图为器件在不同偏压时的能带图

简并半导体类似金属，在绝对零度近似下，费米能级以下的状态全被电子占据，费米分布函数 $f(E)=1$；费米能级以上的状态全空，$f(E)=0$。这样，参考第一章中的公式(1.27)和(1.36)，我们不难得到

$$qV_n = \frac{h^3}{2m_n}\left(\frac{3n}{8\pi}\right)^{2/3} \tag{6.29}$$

$$qV_p = \frac{h^3}{2m_p}\left(\frac{3p}{8\pi}\right)^{2/3} \tag{6.30}$$

m_n和 m_p分别为电子和空穴的有效质量。以 GaAs 为例，$m_n=0.068m_0$(m_0为自由电子质量)，$m_p=0.64m_0$。设 n 区的电子浓度 $n=10^{19}\text{cm}^{-3}$，p 区的空穴浓度 $p=5\times10^{19}\text{cm}^{-3}$，则有 $qV_n\simeq0.24\text{eV}$，$qV_p\simeq0.07\text{eV}$。对于这样的 p-n 结，只是 $V<0.31\text{V}$ 时才有显著的隧道电流，相应的负阻 $R\simeq-3V_p/I_p$。

2. 微波性能

图 6.5 表示隧道二极管偏置在负阻区的小信号等效电路，它由四个元件组成：串联电感 L_s、串联电阻 R_s、二极管电容 C 和二极管负阻 $-R$。C 通常在谷值电压下测量。

等效电路的输入阻抗为

$$Z_{in} = R_s + j\omega L_s + \frac{-R\cdot 1/j\omega C}{-R+1/j\omega C} \tag{6.31}$$

即

$$Z_{in} = \left[R_s - \frac{R}{1+(\omega CR)^2}\right] + j\left[\omega L_s - \frac{\omega CR^2}{1+(\omega CR)^2}\right] \tag{6.32}$$

对于二极管不出现负阻的最高电阻截止频率 f_c，Z_{in}的实部必须是零，故有

$$f_c = \frac{1}{2\pi RC}\sqrt{\frac{R}{R_s} - 1} \tag{6.33}$$

对于自共振频率 f_r, Z_{in}的虚部必须是零,故有

$$f_r = \frac{1}{2\pi RC}\sqrt{\frac{R^2C}{L_s} - 1} \tag{6.34}$$

若 $f_r < f_c$,则二极管将在 f_r 下振荡。

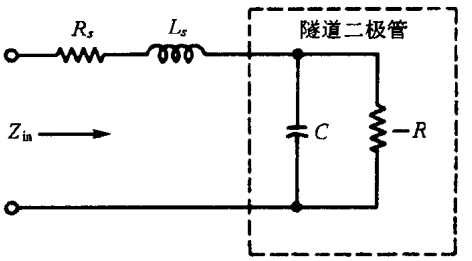


图 6.5 隧道二极管等效电路

峰值电流与谷值电流之比 I_p/I_V 决定最大输出功率和直流转换为交流的效率。为了推导最大输出功率,我们简单假定在负阻区($V_p \leqslant V \leqslant V_V$)电流随电压增加而线性下降;因而对于正弦电压波,最大输出功率为

$$P_{max} = \frac{1}{2\pi}\left(\frac{\Delta I}{2}\right)\left(\frac{\Delta V}{2}\right)\int_0^{2\pi}\sin^2\theta d\theta = \frac{1}{8}(\Delta I \Delta V) \tag{6.35}$$

ΔI 和 ΔV 是负阻区电流和电压的变化范围,$\Delta I = I_p - I_V$, $\Delta V = V_V - V_p$。最大转换效率为

$$\eta = \frac{P_{max}}{P_{dc}} = \frac{\frac{1}{8}(\Delta I \Delta V)}{(I_p - \Delta I/2)(V_p + \Delta V/2)} = \frac{1}{2}\frac{\left(\frac{I_p}{I_V} - 1\right)\left(\frac{V_V}{V_p} - 1\right)}{\left(\frac{I_p}{I_V} + 1\right)\left(\frac{V_V}{V_p} + 1\right)} \tag{6.36}$$

P_{dc}为器件提供的直流功率。由上式可以看出,为了提高最大输出功率和转换效率,必须提高峰谷电流比 I_p/I_V。峰谷比的典型值对 Si 为 4:1, Ge 为 10:1, GaAs 为 12:1。一般说来,对一定的半导体,提高 n 区和 p 区的掺杂浓度可以提高峰谷比;但高的掺杂浓度会导致相当大的结电容,限制截止频率 f_r。归根结底,峰谷比取决于峰值电流和谷值电流,前者随载流子有效质量和禁带隙的减小而增加,后者随带隙内能级浓度的减少而下降。

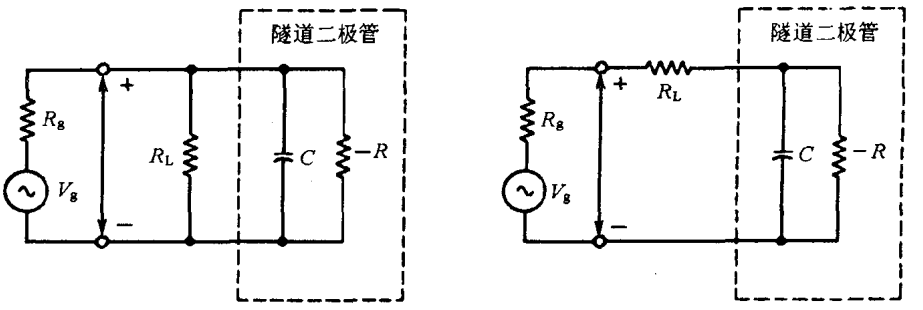
图 6.6 表示隧道二极管并联或串联负载电阻 R_L 组成放大器时的等效电路。由图 6.6(a)可见,放大器的功率增益为

$$G = \frac{R}{R - R_L} \tag{6.37}$$

当隧道二极管的负阻与负载电阻接近时,增益 G 趋于无限大,系统开始振荡。而对于串联负载电阻的情形,由图 6.6(b)可见,功率增益 G 由下式给出:

$$G = \frac{R_L}{R_L - R} = \frac{1}{1 - R/R_L} \tag{6.38}$$

当 $R_L < R$ 时,器件稳定在负阻区,而不会有低阻的导通状态和高阻的关断状态之间进行转换的开关动作。



(a) 并联负载　　(b) 串联负载

图 6.6　隧道二极管放大器等效电路

6.2　共振隧穿二极管(RTD)

将一个双异质结构成的量子阱(双势垒结构)置于两个掺杂的欧姆接触层中间就构成了共振隧穿二极管(RTD)。图 6.7(a)示意画出了 RTD 的一种典型结构。其中,二层厚约 15Å 的 AlAs 材料(禁带隙 2.16eV)夹着一层厚约 50Å 的 GaAs 材料(禁带隙 1.42eV)形成对称的双势垒结构,如图 6.7(b)所示。分别作为势垒层和势阱层的 AlAs 材料和 GaAs 材料一般是不掺杂的。双势垒两侧厚度较大的 n 型 GaAs 层作为欧姆接触和双势垒之间的低阻通道,通常有较高的掺杂浓度。由于具有量子阱结构,比起通常的隧道二极管来,RTD 不仅隧穿几率大而能够提供较大的输出功率,而且结电容小而具有很高的截止频率,所以近年来受到很大重视。

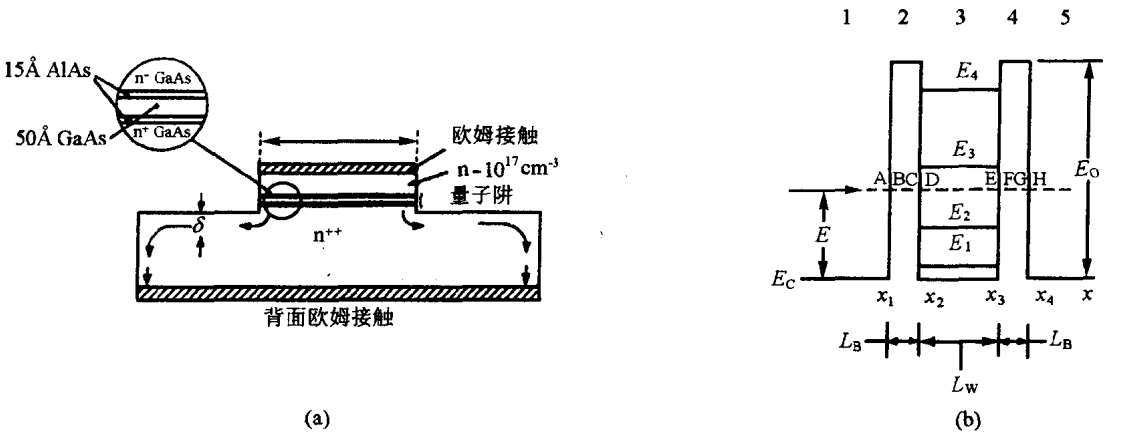


图 6.7　(a) RTD 的典型结构;(b) 一维双势垒结构

6.2.1　共振隧穿效应

图 6.7(b)所示的对称双势垒结构和一个简单势垒的差别,是在其中能够形成一些量子化能级 $E_n(n=1,2,3,\cdots)$。图上只表示出沿 x 方向(垂直于势垒的方向)的运动取量子化的值,实际上应加上与势垒平面平行的 y-z 平面内以二维波矢 $\boldsymbol{k}_\perp$ 表征的自由运动。这样,对于每一

个 E_n，量子阱的允许能级构成一个连续能带，称为子能带(这些子能带部分重叠)。这种势垒结构的特点是，向它入射的电子沿 x 方向运动的能量 E 和阱中的量子化能级(例如 E_1)一致时，有很大的隧穿几率(透射率)，称为共振隧穿(更确切地说，实际三维空间的入射电子还必须和阱中的横向动量 $\hbar k_\perp$ 相同的子带能级共振)。尽管电子对单个势垒的透射率也许不到 1%，但对于这种对称双势垒，其透射率可达 100%；另一方面，当能量稍有偏离时，透射率迅速下降。

电子通过这样一个一维的对称双势垒结构，十分类似于光束通过法布里-珀罗腔(F-P 腔)。在光学 F-P 腔中，从一端进入的光波在腔两端的半透明反射镜之间来回反射，同时相干叠加，结果使得某些波长的光透射率极大，而其他波长的光透射率极小。在双势垒结构中，对电子而言，中间的势阱(量子阱)相当于 F-P 腔，两边的势垒起到腔两端半反射镜的作用。为了理解共振隧穿现象，下面仿照 F-P 干涉仪的原理对电子隧穿双势垒的一维运动进行分析。

设入射波的波幅为 1，r_{ij} 和 t_{ij} 分别代表电子隧穿单个势垒时的反射波和透射波的复振幅，i 代表入射区，j 代表出射区，在图 6.7(b)所示的结构中，设电子从左边入射，连续隧穿两个势垒后从右边射出，出射波的复振幅为 s，可以设想，入射波首先隧射左边的势垒，射入势阱的波的复振幅为 t_{13}，这个波在势阱中来回反射，每在右边的势垒反射一次，都有一部分(t_{35})从右边势垒射出，总的出射波(振幅为 s)可视为多束波隧穿右边势垒的出射波的叠加，即

$$\begin{aligned} s &= s_1 + s_2 + s_3 + \cdots \\ &= t_{13}t_{35} + t_{13}r_{35}r_{31}t_{35} + t_{13}(r_{35}r_{31})^2 t_{35} + \cdots \\ &= t_{13}t_{35}[1 + r_{35}r_{31} + (r_{35}r_{31})^2 + \cdots] \\ &= \frac{t_{13}t_{35}}{1 - r_{35}r_{31}} \end{aligned} \tag{6.39}$$

总的透射率(出射波几率流与入射波几率流之比)为

$$\begin{aligned} T &= \frac{k_5}{k_1} s \cdot s^* \\ &= \frac{k_5}{k_1} \cdot \frac{T_{13}T_{35}}{1 + R_{35}R_{31} - 2\sqrt{R_{35}R_{31}}\cos(\theta_{35} + \theta_{31})} \end{aligned} \tag{6.40}$$

其中，$T_{ij} = t_{ij}t_{ij}^*$，$R_{ij} = r_{ij}r_{ij}^*$，θ_{35} 和 θ_{31} 分别是 r_{35} 和 r_{31} 的相角，它们与势垒的高度和宽度以及势阱宽度有关。在一般情形下，$T \simeq CT_{13}T_{35}$，C 的数量级为 1，此即为相继隧穿过程，两个势垒之间量子阱的存在对隧穿影响不大；但对于电子能量的某些特殊值，将使得 $\theta_{35} + \theta_{31} = 2n\pi$ (n 为整数)，这时透射率有最大值：

$$T_{\max} = \frac{k_5}{k_1} \cdot \frac{T_{13}T_{35}}{(1 - \sqrt{R_{35}R_{31}})^2} \tag{6.41}$$

通常，单个势垒的透射率 T_{ij} 很小，反射率 R_{ij} 很大，因而共振隧穿几率 $T_{\max}$ 比相继隧穿几率 $T \simeq CT_{13}T_{35}$ 大得多。对于图 6.7(b)所示的对称双势垒，$k_5 = k_1$($k = \hbar^2/m$，m 为电子有效质量)，$R_{31} = R_{35}$，$T_{35} = T_{13} = T_{31}$，$R_{ij} + T_{ij} = 1$，式(6.41)给出的 $T_{\max} = 1$，即此时对称双势垒结构对电子完全透明，无反射波。

6.2.2 电流-电压特性

对 RTD 外加电压 V，且极性为左负为正，则可得到如图 6.8 所示的电流-电压特性；该图

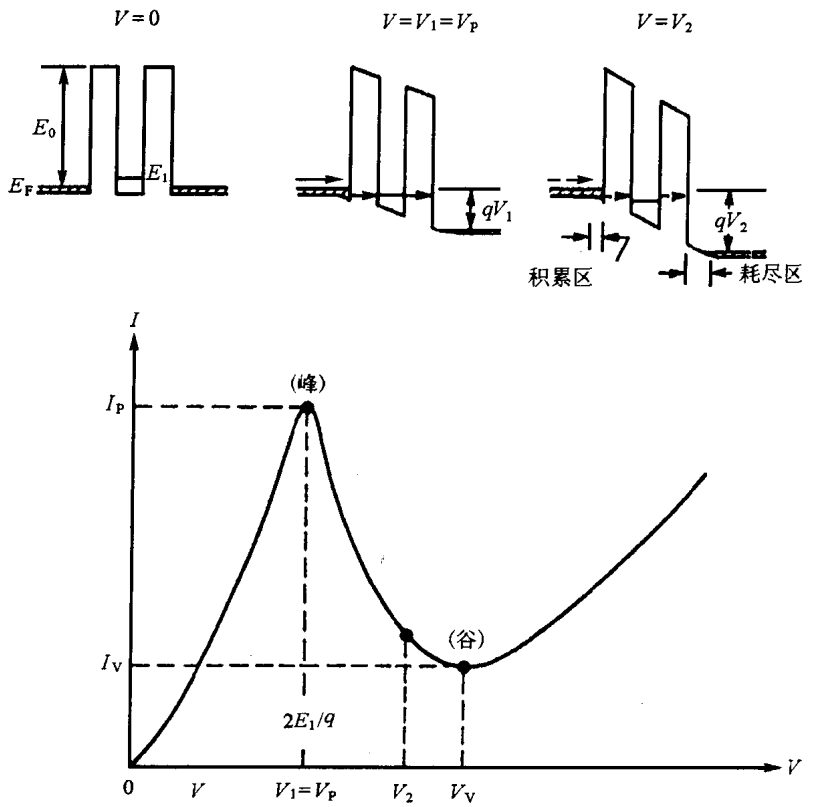


图 6.8　RTD 的电流-电压特性及其在不同直流电压时的能带图

还画出了 RTD 在不同 V 值时的能带。图 6.8 中热平衡($V=0$)时的能带图不同于图 6.7(b)之处是只考虑了一个最低能级 E_1。随外加电压增加,在阴极一侧(发射区)的势垒附近形成积累层,而在阳极一侧(收集区)的势垒附近形成耗尽层。当 V 较小时,只有不多的一点电子可以隧穿通过双势垒,流过器件的电流很小。一旦电压 V 使发射区的 E_F 提高到与 E_1 齐平时,共振发生,费米能级附近能态上的电子可以隧穿左边的势垒进入量子阱,然后隧穿右边的势垒进入收集区导带中没有被占据的状态。随电压进一步增加,有更多的电子参与共振隧穿,电流迅速上升,直至左侧的导带底 E_c 与 E_1 齐平,电流达到峰值 I_p,相应的峰值电压 V_p 满足

$$V_p > \frac{2E_1}{q} + \Delta V_{da} \tag{6.42}$$

ΔV_{da}为积累层和耗尽层上的电压降。当 V 增加至 V_2 时,E_c 上升而超过 E_1,全部电子脱离共振,电流变得很小。V_v 时的谷值电流 I_v 主要来自过剩电流 I_x,包括经由势垒材料高能谷的隧穿过程以及声子或杂质协助(非弹性散射)的隧穿过程,都将导致过剩电流。对于更高的电压($V>V_v$),电子经由量子阱较高分立能级的注入或越过势垒的热离子注入而引起热离子电流 I_{th},这十分类似通常隧道二极管中的热扩散电流。

由于共振隧穿(RT)效应,RTD 和 p-n 结隧道二极管相比,可以有高得多的峰值电流 I_p 和电流峰谷比 I_p/I_v。但是,要获得高的峰谷比和峰值电流,RTD 需要具备几个方面的条件。首先,势阱和两个势垒的总宽度,即 L_W+2L_B,应小于电子的平均自由程 λ(等于电子的热运动

速度 $\sqrt{3k_BT/m}$ 和动量弛豫时间 $m\mu_n/q$ 的乘积，即 $\mu_n\sqrt{3mk_BT}/q$，其中 μ_n 为电子迁移率）；其次，热离子发射电流应比隧道电流小得多。前者要求采用不掺杂的势阱和势垒材料，并改进异质结界面质量以减小杂质和缺陷的散射；后者除了要求势垒高度 E_0（由两种薄层材料的带隙差 ΔE_c 决定）较小和势垒宽度 L_B 较小外，最好在低温下使用以减小热离子发射和降低晶格散射。此外，应采用电子有效质量小的材料，使高掺杂发射区的简并量大，以提高峰值电流。

6.2.3 微波性能

RT 是一种特别快速的量子力学过程，实验和理论模拟都表明响应时间小于 0.3ps，所以相应的时间延迟通常是可以忽略的。在此条件下，RTD 和 p-n 结隧道二极管有相同的小信号等效电路，如图 6.5 所示。但是 RTD 的截止频率 f_r 要高，这是因为：对于 p-n 结隧道二极管，两侧的掺杂浓度十分高（$\gtrsim 10^{19}cm^{-3}$），所以结电容相当大；而对于 RTD，电容主要来自耗尽区（参见图 5.8），那里的掺杂浓度可以做到较小（$\simeq 10^{17}cm^{-3}$），从而耗尽层电容较小。

此外，如果忽略 RT 延迟时间和耗尽层渡越时间，则 RTD 的输出功率和转换效率的表达式和 p-n 结隧道二极管的相同，可分别由式(6.35)和式(6.36)表示，在此不另作讨论。

6.3 IMPATT 二极管

IMPATT(impact ionization avalanche transit time)的含义是碰撞电离雪崩渡越时间。IMPATT 二极管是一种渡越时间二极管，由注入区和漂移区组成，载流子在注入区（雪崩区）内通过强电场下的碰撞电离产生，然后注入漂移区。这样，由两种延迟使电流落后于电压并引起负阻，一种是雪崩电流建立时间所产生的“雪崩延迟”，另一种是载流子通过漂移区时间所产生的“渡越时间延迟”。如果漂移区由两段组成，在其中一段的速度明显比在另一段要小，则由于低速区的渡越时间延迟，可以期望有较高的转换效率；相应的器件称为 DOVATT（双速雪崩渡越时间）二极管。

顺便指出，由于载流子产生和注入的机制不同，还有几类渡越时间二极管。其中，载流子越过势垒（正偏的 p-n 结或肖特基势垒，或异质结）以热离子发射方式注入漂移区的，称为 BARITT（势垒注入渡越时间）二极管；载流子隧穿单势垒（反偏的简并 p-n 结或薄异质结）或共振隧穿双势垒注入漂移区的，称为 TUNNETT（隧穿渡越时间）或 QWITT（量子阱注入渡越时间）二极管；强电场下载流子通过碰撞电离和隧穿效应的共同作用而注入漂移区的，称为 MITATT（混合隧道-雪崩渡越时间）二极管。IMPATT 二极管是最重要的微波有源器件之一，可以在毫米波时产生最大连续功率，被广泛用于 30 至 300GHz 这一频段。其他渡越时间器件，由于具有不同特性，例如噪声低（BARITT 二极管）、效率高（DOVATT 和 QWITT 二极管）或工作频率高（MITATT 和 TUNNETT 二极管），可以弥补 IMPATT 二极管的不足。

本节通过一种特殊掺杂分布的器件（里德二极管）来分析“雪崩延迟效应”和“渡越时间延迟效应”，然后进一步说明 IMPATT 二极管的基本性质。

6.3.1 里德二极管

里德(Read)在 1958 年提出了一种 IMPATT 二极管，在一端有一雪崩区用以产生载流子，在另一端有较高的电阻率（低掺杂浓度）提供渡越空间（漂移区），这就是 p^+-n-i-n^+ 或 n^+-p-i-

p^+结构。图 6.9 表示 p^+-n-i-n^+结构的电场分布和击穿条件下电离率的空间变化。

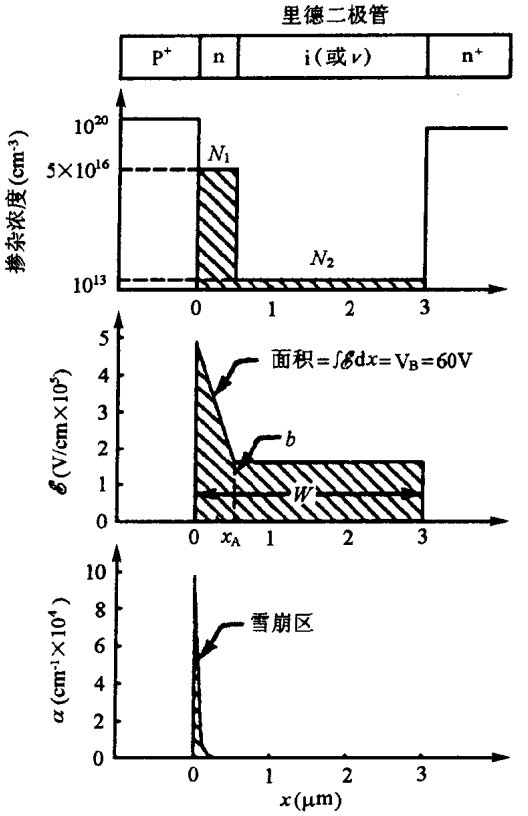


图 6.9　里德型 IMPATT 二极管的掺杂分布、电场分布和电离率的空间变化

在第二章中曾讨论过,雪崩击穿条件为

$$\int_0^W \alpha \mathrm{d}x = 1 \tag{6.43}$$

W 为耗尽区总宽度。从以前的讨论可知,电离率 α 与电场 $\mathscr{E}$ 有强烈的依赖关系,因而它只是在靠近最大电场的 0 和 x_A 之间的窄区域内才明显不为零,称这一区域为雪崩区,其宽度 x_A 比整个耗尽区的宽度 W 小得多。剩余的耗尽区域 $W-x_A$ 为漂移区。这种理想情形如图6.10所示。

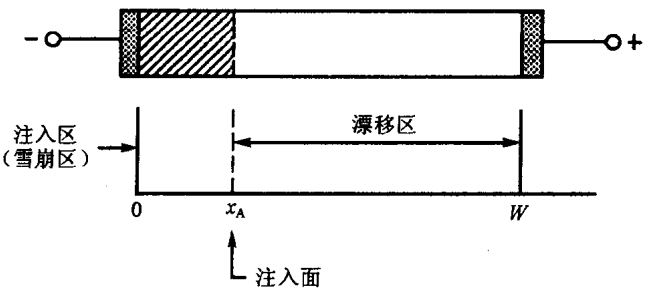


图 6.10　理想 IMPATT 二极管的剖面图

图 6.10 中 $x=x_A$ 为注入面,表示载流子在狭窄的雪崩区(注入区)内产生并在 $x=x_A$ 处注入

邻近的漂移区。

6.3.2 雪崩延迟和渡越时间效应(小信号分析)

根据上述讨论,可以把理想 IMPATT 二极管(里德二极管)分成三个区域:(1) 雪崩区,该区很薄,从而空间电荷和渡越时间效应都可以忽略;(2) 漂移区,该区不产生载流子,从雪崩区进入的载流子全部以饱和速度运动;(3) 无源区,该区加进了不希望有的寄生电阻。

雪崩区 为简化分析,假定电子和空穴的电离率相等($\alpha_n=\alpha_p=\alpha$),并均以相同的饱和速度运动($v_n=v_p=v_s$)。这样,根据第一章的讨论,碰撞电离引起的电子-空穴对产生率为$\alpha(n+p)v_s$,它比热产生率大得多,从而后者可以忽略。于是,有连续方程:

$$\frac{\partial n}{\partial t}=\frac{1}{q}\frac{\partial J_n}{\partial x}+\alpha(n+p)v_s \tag{6.44}$$

$$\frac{\partial p}{\partial t}=-\frac{1}{q}\frac{\partial J_p}{\partial x}+\alpha(n+p)v_s \tag{6.45}$$

将以上二式相加,得到传导电流 J_c 随时间变化的方程:

$$\frac{1}{v_s}\frac{\partial J_c}{\partial t}=\frac{\partial}{\partial x}(J_n-J_p)+2\alpha J_c \tag{6.46}$$

$$J_c=J_n+J_p=q(n+p)v_s \tag{6.47}$$

将方程(6.46)从 $x=0$ 积分到 $x=x_A$,得到

$$\tau_A\frac{dJ_c}{dt}=(J_n-J_p)_0^{x_A}+2J_c\int_0^{x_A}\alpha dx \tag{6.48}$$

其中 $\tau_A=x_A/v_s$。我们注意到,直流传导电流密度(参见第二章)为

$$J_\infty=J_s\left(1-\int_0^{x_A}\alpha dx\right)^{-1} \tag{6.49}$$

J_s 为反向饱和电流密度。为了使方程(6.48)在 $dJ_c/dt=0$ 的情形下简化为式(6.49),必须有

$$(J_n-J_p)_0^{x_A}=2(J_s-J_c) \tag{6.50}$$

换句话说,方程(6.48)应具有下述形式(忽略 J_s 项):

$$\frac{dJ_c}{dt}=\frac{2J_c}{\tau_A}(\bar{\alpha}x_A-1) \tag{6.51}$$

$$\bar{\alpha}x_A-1=\int_0^{x_A}\alpha dx \tag{6.52}$$

$\bar{\alpha}$ 是 α 的空间平均,通过在整个雪崩区上计算积分得到。在既有直流又有交流的情形下,可假定

$$\bar{\alpha}=\bar{\alpha}_0+\overline{\frac{d\alpha}{d\mathscr{E}}}\Delta\mathscr{E}=\bar{\alpha}_0+\bar{\alpha}'\mathscr{E}_1^A \tag{6.53}$$

$\mathscr{E}_1^A$ 是雪崩区电场的交流分量。以脚标"0"和"1"分别代表直流分量和交流分量,则由方程(6.51)可得

$$J_{c1}=\frac{2J_{c0}\bar{\alpha}'x_A}{j\omega\tau_A}\mathscr{E}_1^A \tag{6.54}$$

得到上式时利用了 $J_{c1}(\bar{\alpha}_0x_A-1)$ 可以忽略的条件,这是因为在雪崩条件下 $\bar{\alpha}_0x_A$ 接近于 1。

在交流情形下,总电流中还应包括位移电流。流过雪崩区的位移电流为

$$J_d = j\omega\varepsilon_s \mathscr{E}_1^A \tag{6.55}$$

ε_s 为半导体的介电常数。将式(6.54)和(6.55)相加,得到总的交流密度为

$$J_1 = J_{c1} + J_d = \left[\frac{2J_{c0}\overline{\alpha}' x_A}{j\omega\tau_A} + j\omega\varepsilon_s\right]\mathscr{E}_1^A \tag{6.56}$$

流过 IMPATT 二极管的交流电流为 $J_1 A$(A 是二极管的横截面积),而雪崩区上的交流电压为 $\mathscr{E}_1^A x_A$。由式(6.56)不难看出,交流传导电流比 $\mathscr{E}_1^A$(交流电压)落后 $\pi/2$,如同在电感器内一样;位移电流则超前 $\pi/2$,如同在电容器内一样。所以雪崩区的作用可以用一个并联的 LC 四路等效,其中

$$L_A = \frac{\tau_A}{2J_{c0}\overline{\alpha}' A} \quad , \quad C_A = \frac{\varepsilon_s A}{x_A} \tag{6.57}$$

等效电路的共振频率 ω_r 可表示为

$$\omega_r^2 = \frac{2J_{c0}\overline{\alpha}' v_s}{\varepsilon_s} \tag{6.58}$$

ω_r 称为雪崩谐振频率。利用 ω_r 可以将 J_1 表示成

$$J_1 = \left[1 - \frac{\omega_r^2}{\omega^2}\right] j\omega\varepsilon_s \mathscr{E}_1^A \tag{6.59}$$

不难看出雪崩区的阻抗为

$$Z_A = \frac{1}{j\omega C_A}\left(1 - \frac{\omega_r^2}{\omega^2}\right)^{-1} \tag{6.60}$$

交流传导电流 J_{c1} 在总的交流电流 J_1 中所占比例为

$$\gamma = \frac{J_{c1}}{J} = \frac{1}{1 - (\omega/\omega_r)^2} \tag{6.61}$$

漂移区 如果把雪崩区和漂移区的交界面(注入面)取作 $x = 0$,则离开雪崩区的电流要经过 $t' = x/v_s$ 才能到达漂移区中的 x 处,所以 t 时刻 x 处的交流传导电流 $J_{c1}(x)$ 就是 $t - t'$ 时刻 $x = 0$ 处的传导电流,故有

$$J_{c1}(x) = \gamma J_1 e^{-j\omega x/v_s} \tag{6.62}$$

漂移区中的位移电流也是位置 x 的函数,可表示为

$$\begin{aligned} J_d(x) &= J_1 - J_{c1}(x) = J_1\left[1 - \gamma e^{-j\omega x/v_s}\right] \\ &= j\omega\varepsilon_s \mathscr{E}_1^D(x) \end{aligned} \tag{6.63}$$

$\mathscr{E}_1^D(x)$ 是漂移区中的交流电场,由上式得到为

$$\mathscr{E}_1^D(x) = \frac{J_1}{j\omega\varepsilon_s}\left[1 - \gamma e^{-j\omega x/v_s}\right] \tag{6.64}$$

将式(6.64)从 $x = 0$ 积分到 $x = W - x_A$,就得到漂移区上的交流电压:

$$\begin{aligned} V_1^D &= \frac{(W - x_A)}{j\omega\varepsilon_s} J_1\left[1 - \gamma\frac{1 - e^{j\theta}}{j\theta}\right] \\ &= \frac{(W - x_A)}{j\omega\varepsilon_s} J_1\left[1 - \frac{\gamma\sin\theta}{\theta} - \frac{\gamma(1 - \cos\theta)}{j\theta}\right] \end{aligned} \tag{6.65}$$

其中 θ 为漂移区的渡越角:

$$\theta = \omega(W - x_A)/v_s \tag{6.66}$$

漂移区的阻抗为

$$Z_D = \frac{V_1^D}{J_1 A} = \frac{W - x_A}{j\omega\varepsilon_s A}\left[1 - \frac{\gamma\sin\theta}{\theta} - \frac{\gamma(1-\cos\theta)}{j\theta}\right]$$
$$= R_D + jX_D \tag{6.67}$$

等效电阻为

$$R_D = \frac{(W - x_A)}{\omega\varepsilon_s A} \cdot \frac{\gamma(1-\cos\theta)}{\theta} = \left[\frac{1}{1 - \omega^2/\omega_r^2}\right]\left[\frac{1-\cos\theta}{\omega C_D \theta}\right] \tag{6.68}$$

其中 $C_D = \varepsilon_s A/(W - x_A)$为漂移区的电容。$\omega > \omega_r$ 时，R_D 出现负值，$\theta = \omega(W - x_A)/v_s = \omega\tau_D$ 较大(即渡越时间 τ_D 足够长)时这一负值才显著不为零。对于 Si，对应 $\mathscr{E} = 3.8 \times 10^5 \text{V/cm}$，$\overline{\alpha'} \simeq 0.25/\text{V}$ 和 $v_s \simeq 8 \times 10^6 \text{cm/s}$；若 $J_{c0} = 1\ 300 \text{A/cm}^2$，则 $\omega_r = 0.7 \times 10^{11} \text{rad/s}$，属于微波波段。

图 6.11 是根据上述小信号分析得到的 IMPATT 二极管等效电路，器件的总阻抗包括雪崩区和漂移区的阻抗以及无源区的寄生电阻 R_s。雪崩区的阻抗是电抗性的，所以器件的总电阻为

$$R = R_D + R_S$$
$$= \left[\frac{1}{1 - \omega^2/\omega_r^2}\right]\left[\frac{1-\cos\theta}{\omega C_D \theta}\right] + R_S \tag{6.69}$$

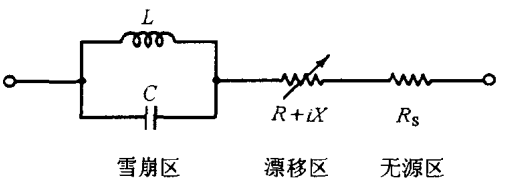


图 6.11　IMPATT 二极管的小信号等效电路

6.3.3　功率和效率(大信号分析)

如前所述，对于 p^+-n-i-n^+ 结构的 IMPATT 二极管，在产生电子-空穴对的 p^+-n 结处(图 6.12(a))有一高场雪崩区，而近似本征半导体的轻掺杂区(i 区或 ν 区)是电场恒定的漂移区。所产生的空穴迅速进入 p^+ 区，所产生的电子则注入漂移区。在漂移区，电子作功，产生外功率。当电场围绕平均值随时间周期性变化时(图 6.12(c))，每个载流子的碰撞电离率几乎瞬时地跟随电场变化；然而，由于载流子的产生还依赖于已经存在的载流子数，载流子浓度不能同步跟随电场变化。电场通过峰值以后，由于载流子的产生率仍然高于平均值，故载流子浓度持续增加。当电场从峰值减小到平均值时，载流子浓度近似达到最大值。因此，即使电离率与交流电场同相，注入载流子浓度的交流变化仍滞后于交流电压约 π/2(即所谓"雪崩延迟")。这就是图 6.12(d)中的"注入电流"的情形。交流电场(或电压)的峰值出现于 π/2 处，而注入载流子浓度("注入电流")的峰值却出现于 π 处(即 $\phi = \pi$)。注入载流子(电子)一进入漂移区就会在外电路中感生电流，直到它们全部通过漂移区为止，这种情形相当于正电荷从 p^+ 端源源不断地流入 n^+ 端，当电子全部到达 n^+ 端时，正电荷也全部到达，正负电荷在这里中和。所以，对于注入电子以恒定速度(饱和漂移速度 v_s)渡越漂移区的情形，感生的外电流近似等于

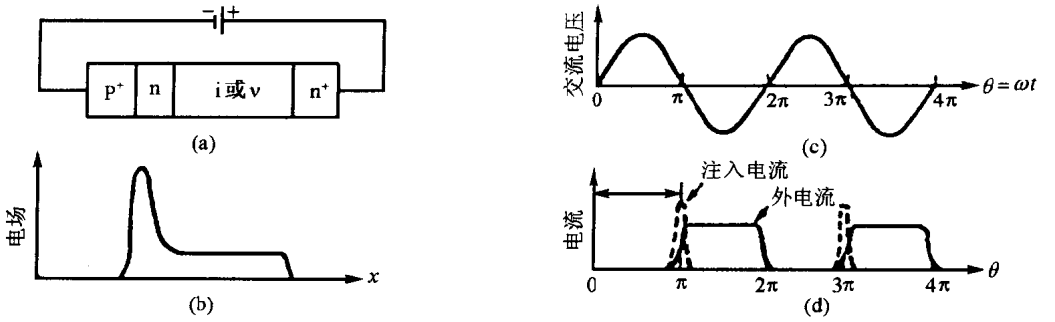


图 6.12 (a) p^+-n-i-n^+结构；(b) 雪崩击穿时的电场；(c) 交流电压；(d) 注入电流和外电流

注入电荷脉冲Q除以渡越时间τ_D，即

$$I_{ind} \simeq Q/\tau_D \tag{6.70}$$

感生的外电流也表示在图 6.12(d)中。对交流电压和外电流进行比较后可以清楚地看出，IMPATT 二极管表现出负阻。

在理想情形下，各类渡越时间二极管总是穿通，其漂移区中的电场总是高到载流子足以保持饱和速度 v_s。这一共同特点使我们可以用一个统一的大信号分析模型来描述它们的功率特性。图 6.13 表示渡越时间二极管理想的电压和电流波形，其中 δ 是注入电流脉冲的宽度，ϕ 是注入相位延迟而位于脉宽中央。注入相位 ϕ 取决于渡越时间二极管注入机制的性质，例如，对 IMPATT 二极管，$\phi=\pi$；对 BARITT 二极管，$\phi=\pi/2$。

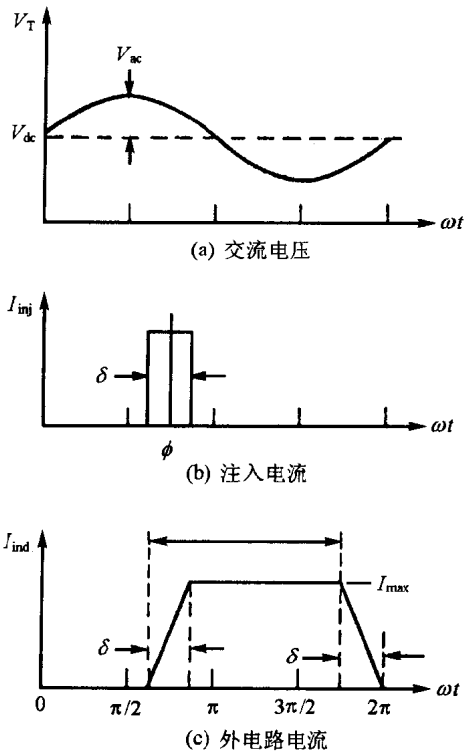


图 6.13 渡越时间二极管电压和电流的理想波形

二极管上的电压由下式给出：

$$V = V_{dc} + V_{ac}\sin\omega t \tag{6.71}$$

其中 V_{dc}, V_{ac}和 ω 分别代表直流电压、交流电压振幅和工作频率。直流电流为

$$I_{dc} = \frac{1}{2\pi}\int_0^{2\pi} I_{ind}\mathrm{d}(\omega t) \tag{6.72}$$

I_{ind}是感生外电流，从图 6.13(c)不难看出其最大值为

$$I_{max} = \frac{2\pi}{\theta} I_{dc} \tag{6.73}$$

直流输入功率为

$$P_{dc} = V_{dc} I_{dc} \tag{6.74}$$

交流输出功率为

$$P_{ac} = \frac{1}{2\pi}\int_0^{2\pi} I_{ind}(\omega t)\, V_{ac}\sin\omega t\, \mathrm{d}(\omega t) \tag{6.75}$$

直流转换为交流的效率 η 可以从方程(6.74)和(6.75)得到。利用图 6.13(c)，写出 I_{ind}作为 ωt 的函数的表达式，并完成方程(6.75)的积分，不难证明，

$$\eta = \frac{P_{ac}}{P_{dc}} = \frac{V_{ac}}{V_{dc}} \cdot \frac{\sin(\delta/2)}{\delta/2} \cdot \frac{\cos\phi - \cos(\phi + \theta)}{\theta} \tag{6.76}$$

对于 IMPATT 二极管，$\phi = \pi$，方程(6.76)简化为

$$\eta = \frac{V_{ac}}{V_{dc}} \cdot \frac{\cos\theta - 1}{\theta} \cdot \frac{\sin(\delta/2)}{\delta/2} \tag{6.77}$$

为了提高 η，应当减小注入脉宽 δ，若 θ 一定，其最佳效率可在尖脉冲(即 $\delta \simeq 0$)条件下获得，

$$\eta = \frac{V_{ac}}{V_{dc}} \frac{\cos\theta - 1}{\theta} \tag{6.78}$$

当 $\theta = 0.74\pi$ 时得到最大效率，为

$$\eta = \frac{2.27}{\pi} \frac{V_{ac}}{V_{dc}} \tag{6.79}$$

若 $V_{ac}/V_{dc} = 0.5$，则最大效率约为 36%。

输出功率限制 单个 IMPATT 二极管的最大输出功率 P_m 受雪崩击穿电压 V_m(或临界电场 $\mathscr{E}_c$)和饱和漂移速度 v_s 的限制。对于均匀雪崩，$V_m = \mathscr{E}_c W$(W 为耗尽区宽度)，通过二极管的最大电流则为 $I_m = qnv_sA$(A 为横截面积)，因此有

$$P_m = V_m I_m = (\mathscr{E}_c W)(qnv_s A) \tag{6.80}$$

能参与导电的载流子数 n 通过泊松方程受空间电荷限制：

$$\frac{\mathrm{d}\mathscr{E}}{\mathrm{d}x} \simeq \frac{\mathscr{E}_c}{W} = \frac{qn}{\epsilon_s} \tag{6.81}$$

将式(6.81)代入(6.80)，可得

$$P_m = \epsilon_s \mathscr{E}_c^2 v_s A \tag{6.82}$$

利用二极管电容 $C = \epsilon_s A/W$ 及其特征频率 $f = v_s/2W$(由漂移角 $\omega\tau_D = \pi$ 得到)，上式又可表成

$$P_m = \frac{\mathscr{E}_c^2 v_s^2}{4\pi f^2 \chi_c} \tag{6.83}$$

其中 $\chi_c = 2\pi fc$ 为器件的电抗。若使 χ_c 限制在某一最小值，则式(6.83)表示 IMPATT 二极管能够给出的最大功率随 $1/f^2$ 下降。在高频下，这个电子学限制是主要的。

IMPATT 二极管的输出功率还受到散热过程的限制。在稳定情形下，二极管能够耗散的功率 P 等于能传输到散热器的功率，即

$$P = \Delta T / R_T \tag{6.84}$$

ΔT 为 p-n 结与散热器之间的温度差，R_T 为总热阻。假定 R_T 主要来自半导体，则 $R_T \simeq W/2A\sigma_s$，σ_s(单位：W/cm·K)为半导体的热导率。如果二极管的电抗 $\chi_c = 2\pi fc$ 保持恒定，则有

$$Pf \sim \frac{\sigma_s \Delta T}{\varepsilon_s} = \text{常数} \tag{6.85}$$

因此，在低频下，连续工作的输出功率随 $1/f$ 下降。

6.4 转移电子器件

转移电子器件(TED)利用强电场下电子在半导体导带的两个不等价能谷之间转移而出现负微分迁移率进行工作。在许多半导体中都观察到这种电子转移效应，其中 n 型 GaAs 和 n 型 InP 得到了最广泛的研究和应用，TED 几乎都用它们制作。不像前述的几类微波有源二极管利用结的界面性质，TED 利用了一些特殊半导体的体材料性质，所以有时也称为体效应器件。TED 和 IMPATT 二极管是两种最重要的微波器件，特别是在毫米波应用方面，与 IMPATT 二极管相比，TED 的输出功率和转换效率要低一些，但具有噪声小、工作电压低和电路设计比较容易等优点。

6.4.1 电子转移和负微分迁移率

所谓电子转移，指导带电子从迁移率高的能谷向迁移率低但能量位置较高的能谷转移。具有这种电子转移性质的半导体通常称为两能谷半导体。例如 GaAs 和 InP，它们的导带底在 Γ 点($k=0$)，而沿〈111〉轴接近 L 点还有能量稍高但有效质量大得多(即电子迁移率小得多)的 8 个等价能谷。对于这两种半导体，L 点和 Γ 点之间的能量差大约分别为 0.31eV 和 0.69eV；在强电场下获得足够高的能量时，电子可以由 Γ 能谷(主能谷)向 L 能谷(次能谷)转移，产生电子转移效应。

在图 6.14 所示的简化能带结构中，电子起初处在主能谷内；在外加电场 $\mathscr{E}$ 下，它们将获得能量。若 $\mathscr{E} < \mathscr{E}_T$($\mathscr{E}_T$ 称为阈值电场)，大多数电子仍然留在主能谷内。当 $\mathscr{E} > \mathscr{E}_T$ 时，电子从电场获得足够能量，许多电子被散射("转移")进入次能谷；次能谷较大的有效质量使得迁移率较小，从而电子的平均漂移速度 v 减小。如图 6.15 所示。能量很高(即电场很强，$\mathscr{E} \gg \mathscr{E}_T$)时，几乎全部电子都转移到次能谷内，平均漂移速度在经过极小值以后重新随电场增加。$\mathscr{E} > \mathscr{E}_T$ 时平均漂移速度下降产生一个负微分迁移率的区域。次能谷的电子在外电场 $\mathscr{E}$ 下降到低于 $\mathscr{E}_T$ 时，将会因能量减小经过散射返回主能谷。

概括地讲，能产生负微分迁移率(负微分电阻)的半导体必须至少有两个不等价的导带能谷，并且必须满足下述基本条件：(1) 晶格温度足够低，或热能 $k_B T$ 小于两个能谷之间的能量差 ΔE，使得热平衡(无偏置电场)时，电子基本上都处在低能谷(主能谷)内；(2) 在低能谷内，电子的有效质量小，迁移率高和态密度低；而在高能谷(次能谷)内，电子的有效质量大、迁移率

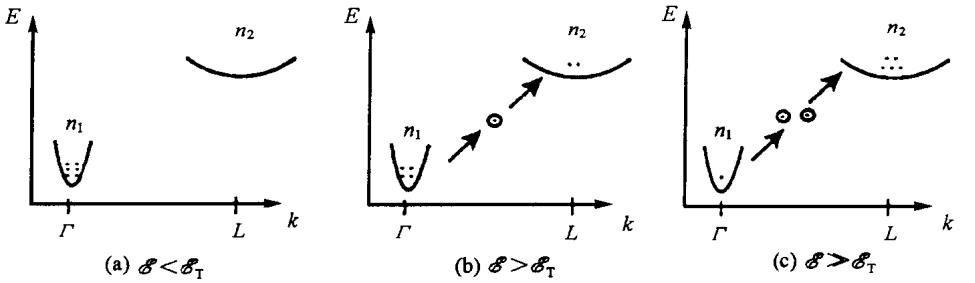


图 6.14　说明电子转移的两能谷直接半导体的简化能带图

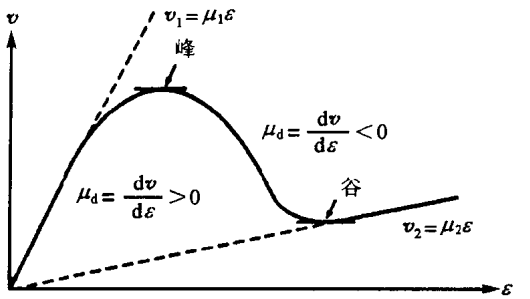


图 6.15　图 6.14 的两能谷半导体的速度-电场分布

低和态密度高;(3) ΔE 小于半导体的禁带宽度 E_g,使得电子在转移到高能谷之前不致发生雪崩击穿。

在满足这些要求的半导体中,n 型 GaAs 和 n 型 InP 的研究和应用最为广泛。图 6.16 表示了室温下这两种半导体实测的电子速度-电场特性。GaAs 的阈值电场为 3.2kV/cm, InP 的为 10.5kV/cm。对 GaAs,峰值速度 v_p 为 2.2×10^7cm/s, Inp 的为 2.5×10^7cm/s。GaAs 的最大负微分迁移率(即 $dv/d\mathscr{E}$)约为 $-2\ 400$cm^2/V·s, Inp 则约为 $-2\ 000$cm^2/V·s。

出现负微分电阻(NDR)的半导体本质上是不稳定的,因为在半导体内任一点的载流子浓度的随机涨落均会产生瞬时空间电荷,这瞬时空间电荷将随时间指数增长。设电子浓度偏离均匀的平衡浓度 n_0(半导体均匀掺杂 N_D)有一小的局部涨落,则产生的局部空间电荷密度为 $q(n-n_0)$,一维连续方程和泊松方程分别为

$$\frac{\partial(n-n_0)}{\partial t}+\frac{1}{q}\frac{\partial J}{\partial x}=0 \tag{6.86}$$

和

$$\frac{\partial\mathscr{E}}{\partial x}=\frac{q(n-n_0)}{\varepsilon_s} \tag{6.87}$$

写出以上二式时,已假定电子有正电荷,以避免多余的负号。此外,为了避免数学上的烦琐而突出空间电荷密度随时间的变化规律,简单假定电流密度为 $J(x)=\sigma\mathscr{E}(x)$($\sigma$ 为电导率),将其对 x 微商并代入泊松方程,则由方程(6.86)可得

$$\frac{\partial(n-n_0)}{\partial t}+\frac{\sigma}{\varepsilon_s}(n-n_0)=0 \tag{6.88}$$

方程(6.88)的解为

$$n - n_0 = (n - n_0)_{t=0} e^{-t/\tau_R} \tag{6.89}$$

式中，τ_R 为介质弛豫时间，用下式表示

$$\tau_R \equiv \frac{\varepsilon_s}{\sigma} \simeq \frac{\varepsilon_s}{qn_0\overline{\mu}} \tag{6.90}$$

$\overline{\mu} = \mathrm{d}v/\mathrm{d}\mathscr{E}$ 为微分迁移率(或平均迁移率)。如果 $\overline{\mu}$ 是正的，τ_R 代表空间电荷恢复到电中性的时间常数。但如果半导体表现出负微分迁移率，则任何不平衡电荷都会按时间常数指数增长。

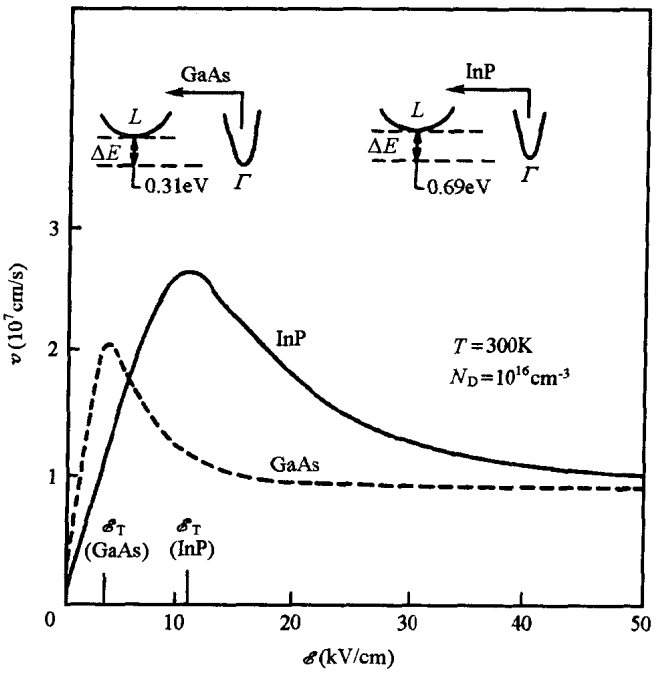


图 6.16　GaAs 和 InP 的速度-电场特性，插图表示电子转移的两能谷模型

6.4.2　偶极畴和基本工作原理

转移电子器件需要高纯度的均匀材料，并使它的深能级杂质和缺陷减少到最低限度。现在的转移电子器件几乎都用各种外延技术在 n^+ 衬底上生长 n 型外延层，并用 n^+ 欧姆接触作为阴极。对于这样的欧姆接触，靠近阴极总有一个低场区，且沿器件的长度方向电场是不均匀的，结果在阴极处形成空间电荷区。

我们已经说明，具有负微分迁移率的器件，初始空间电荷将随时间指数增长[式(6.89)]，其时间常数由式(6.90)给出：

$$\tau_R = \frac{\varepsilon_s}{qn_0|\mu_-|} \tag{6.91}$$

式中 μ_- 是负微分迁移率。如果式(6.89)在整个空间电荷层渡越时间内始终是正确的，则最大增长因子将是 $\exp(L/v|\tau_R|)$，其中 L 是器件长度，v 是空间电荷层平均漂移速度。要使空间电荷大量增长，增长因子必须明显大于 1，这就要使 $L/v|\tau_R| > 1$，或

$$n_0 L > \frac{\varepsilon_s v}{q|\mu_-|} \tag{6.92}$$

对于 n 型的 GaAs 和 InP，上式右方约为 10^{12}cm^{-2}。乘积 n_0L 小于 10^{12}cm^{-2}的转移电子器件呈现稳定的电场分布。所以载流子浓度和器件长度的乘积 $n_0L = 10^{12}\text{cm}^{-2}$是区分转移电子器件各种工作状态的一个重要界限。

空间电荷不稳定性的最简单形式是存在一个积累层。当加一恒定电压时，轻掺杂或短样品（$n_0L < 10^{12}\text{cm}^{-2}$）呈现稳定的电场分布。但在 $n_0L > 10^{12}\text{cm}^{-2}$时，器件中将会形成一个移动的偶极层（也叫畴），其中积累层和耗尽层束缚在一起。图 6.17 表示均匀条件下这种偶极畴的电子浓度和电场分布。电子在低场区（电场 $\mathscr{E}_1$ 一定）以恒定速度 v_T 运动，在区域“a”受到较强的电场加速，直至在区域“b”转移到高能谷使速度减慢而陷入这一积累区内。在区域“c”，电子因能量减小返回低能谷，它们的平均速度比在区域“b”高，从而区域“c”耗尽电子。偶极畴在阴极形成并通过有源区生长和传播，畴的电压增加，畴外电压则在偏置电压一定时下降，即畴外的电场 $\mathscr{E}_1$ 下降，这将阻止有源区内形成新畴，也限制现行畴的增长（这时几乎没有电子陷入积累层或逃离耗尽层）。偶极畴到达阳极时消失，外电路中的电流增加。电流和电压之间的位相差引起动态负阻，在适当的电路中将会产生微波功率。

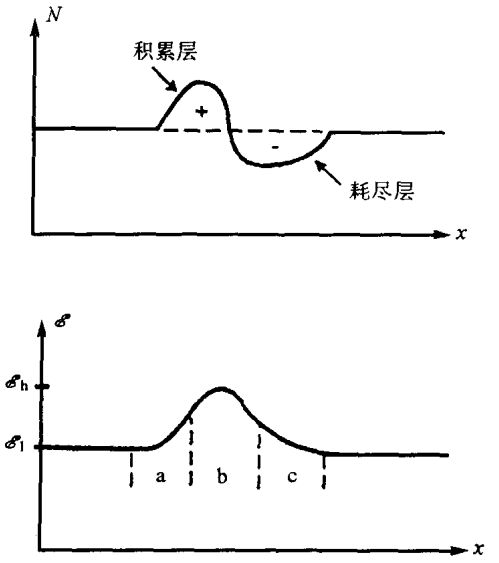


图 6.17　偶极畴的电子浓度和电场分布

我们在前面所作的介质弛豫时间分析只能用来估计初始阶段空间电荷随时间的变化，对于最终将达到稳定状态的偶极畴，这种分析方法是不精确的。下面我们将比较详细地叙述偶极畴的理论。首先，按照前面的约定，将泊松方程表示为

$$\frac{\partial \mathscr{E}}{\partial x} = \frac{q(n - n_0)}{\varepsilon_s} \tag{6.93}$$

其次，对偶极层采用一个简化模型：在积累区 $n \to \infty$，在耗尽区 $n \to 0$，即偶极畴为三角形，畴外电场为 $\mathscr{E}_1$，畴内最大电场为 $\mathscr{E}_h$。在 GaAs 和 InP TED 中，实验得到的只是三角形畴。采用这一简化模型以后，总电流中不出现扩散项，其密度为

$$J = qnv(\mathscr{E}) + \varepsilon_s \frac{\partial \mathscr{E}}{\partial t} \tag{6.94}$$

在畴外，$n = n_0$，$\mathscr{E} = \mathscr{E}_1$，故有

$$J = qn_0 v(\mathscr{E}_1) + \varepsilon_s \frac{\partial \mathscr{E}_1}{\partial t} \tag{6.95}$$

在畴内，利用方程(6.93)可得

$$n = n_0 + \frac{\varepsilon_s}{q} \frac{\partial \mathscr{E}}{\partial x} \tag{6.96}$$

将其代入方程(6.94)，有

$$J = qn_0 v(\mathscr{E}) + \varepsilon_s v(\mathscr{E}) \frac{\partial \mathscr{E}}{\partial x} + \varepsilon_s \frac{\partial \mathscr{E}}{\partial t} \tag{6.97}$$

由于 J 与 x 无关，我们可以从方程(6.95)和(6.97)得到

$$\frac{\partial}{\partial t}(\mathscr{E} - \mathscr{E}_1) = \frac{qn_0}{\varepsilon_s}[v(\mathscr{E}_1) - v(\mathscr{E})] - v(\mathscr{E}) \frac{\partial \mathscr{E}}{\partial x} \tag{6.98}$$

将方程(6.98)在偶极畴上积分，并注意畴外两侧的电场和速度均与 x 无关，从而方程右方第二项对积分结果没有贡献。所以有

$$\frac{\mathrm{d}}{\mathrm{d}t} \int_{x_1}^{x_2} (\mathscr{E} - \mathscr{E}_1) \mathrm{d}x = \frac{qn_0}{\varepsilon_s} \int_{x_1}^{x_2} [v(\mathscr{E}_1) - v(\mathscr{E})] \mathrm{d}x \tag{6.99}$$

x_1 和 x_2 表示偶极畴的范围。

以 V_{ex}表示畴外电场为 $\mathscr{E}_1$ 时偶极畴所含的过剩电压，则

$$V_{ex} = \int_{x_1}^{x_2} (\mathscr{E} - \mathscr{E}_1) \mathrm{d}x \tag{6.100}$$

显然，方程(6.99)左边表示过剩电压随时间的变化率，即偶极畴的生长速率。对于我们采用的三角形畴模型，V_{ex}基本上来自耗尽层($n = 0$)的贡献，因此方程(6.99)右边的积分变量 x 可以通过下式变换为 $\mathscr{E}$：

$$\frac{\partial \mathscr{E}}{\partial x} = -\frac{qn_0}{\varepsilon_s} \tag{6.101}$$

于是，最终可以把方程(6.99)改写为

$$\frac{\mathrm{d}V_{ex}}{\mathrm{d}t} = \int_{\mathscr{E}_1}^{\mathscr{E}_m} [v(\mathscr{E}) - v(\mathscr{E}_1)] \mathrm{d}\mathscr{E} \tag{6.102}$$

方程(6.102)表明，可以用速度-电场(v-$\mathscr{E}$)特性曲线来预测偶极畴的生长速率。当畴内最大电场 $\mathscr{E}_m$ 达到峰值 $\mathscr{E}_h$ 时，偶极畴达到稳定，这时有

$$\int_{\mathscr{E}_1}^{\mathscr{E}_h} [v(\mathscr{E}) - v(\mathscr{E}_1)] \mathrm{d}\mathscr{E} = 0 \tag{6.103}$$

如图 6.18 所示。由该图可见，只有两个阴影区的面积相等时，方程(6.103)才能得到满足，通常称此为“等面积法则”。采用这一法则以后，若畴外电场 $\mathscr{E}_1$ 已知，则可确定畴的峰值电场 $\mathscr{E}_h$。

稳定时，偶极畴的过剩电压为

$$V_{ex} = \frac{\varepsilon_s}{2qN_D} (\mathscr{E}_h - \mathscr{E}_1)^2 \tag{6.104}$$

其中 N_D 是器件有源区的施主掺杂浓度，$\mathscr{E}_h$ 和 $\mathscr{E}_1$ 的关系由等面积法则给出，所以当给定半导

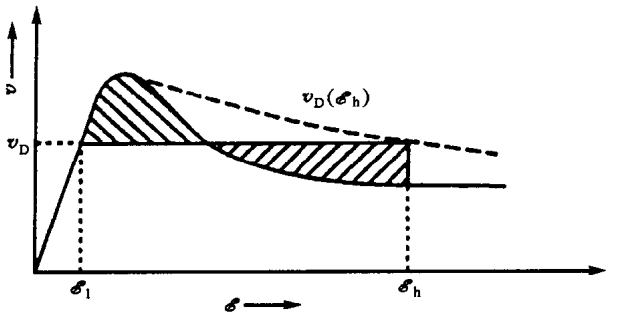


图 6.18　转移电子器件的等面积法则

体的掺杂浓度N_D时，可以得到 V_{ex}和 $\mathscr{E}_1$ 的关系，如图 6.19 中的实线所示，图中不同实线对应不同的掺杂浓度。另一方面，过剩畴电压还必须同时满足下式：

$$V_{ex} = V - L\mathscr{E}_1 \tag{6.105}$$

其中 L 为器件长度，V 为外加偏压。所以选定 L 和 V 时，可以得到图 6.19 中的虚线（器件线），该器件线对应 25μm 的样品长度和 10V 的外加偏压。虚线和实线的交点决定长度为 L，掺杂浓度为 N_D 的器件当外加偏压为 V 时的畴外电场 $\mathscr{E}_1$，这也就决定了畴运动的速度 v_D（等于 $v(\mathscr{E}_1)$）。

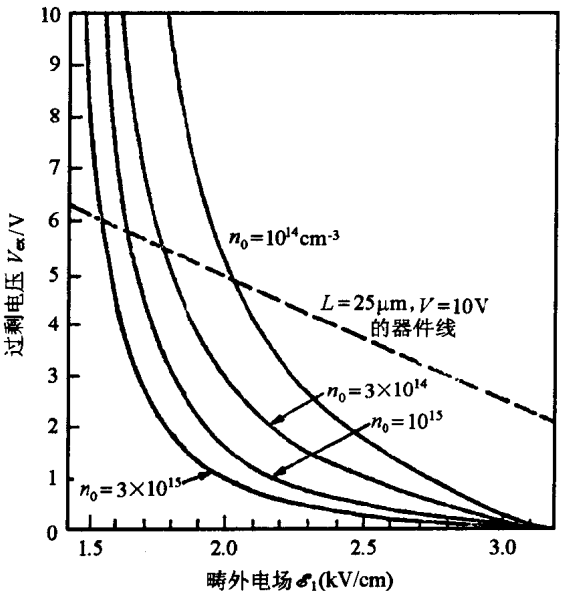


图 6.19　各种掺杂浓度下过剩畴电压和畴外电场的关系

在稳态畴穿越器件的过程中，器件电流由畴外的电场 $\mathscr{E}_1$（即 $v(\mathscr{E}_1)$或 v_D）决定：

$$J = qn_0v_D \tag{6.106}$$

当偶极畴到达阳极时消失，外电路中的电流增加，器件内的电场重新调整，并产生新的畴核。于是，电流振荡的频率依赖于畴速 v_D，可近似表示为

$$f = v_D/L \tag{6.107}$$

L 在微米范围，f 的典型值在 GHz 范围。

6.4.3 器件工作状态

取决于器件本身的性质和外电路情况，TED 有不同的工作模式，其中有三种是最重要的：积累层模式，偶极畴渡越时间模式，猝灭畴模式。

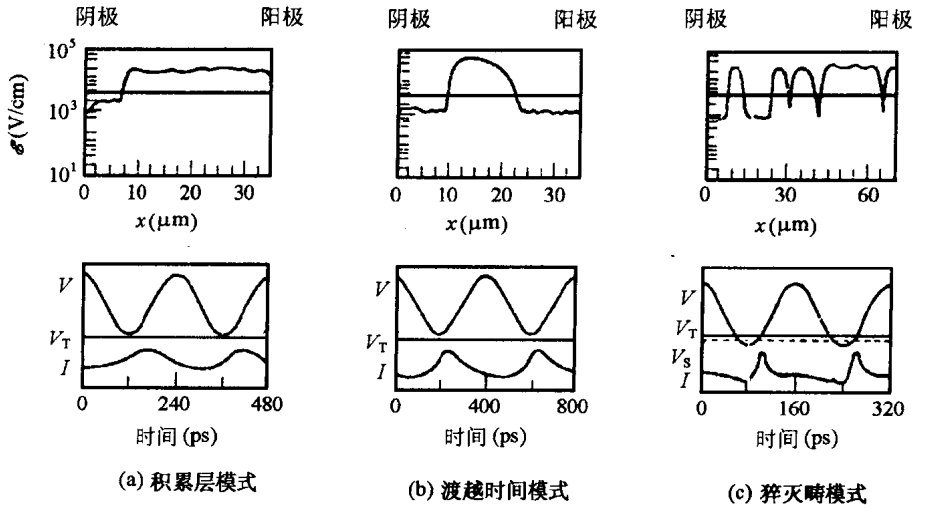


图 6.20 TED 各种工作模式的电场分布、电流和电压波形

积累层模式 具有亚临界 n_0L 乘积(即 $n_0L<10^{12}\text{cm}^{-2}$)的 TED 中，空间电荷没有足够的时间生长或者生长速度太慢。这样的 TED 接到谐振电路上时，如果电压高于阈值($V>V_T=\mathscr{E}_TL$)，将以积累层模式振荡。图 6.20(a)表示射频电压的峰值附近电场与距离的关系，以及电压和电流的波形。这些波形与理想波形偏离很远，直流到交流的转换效率对 GaAs 约为 5%，对 InP 约为 7.5%。

偶极畴渡越时间模式 当 $n_0L>10^{12}\text{cm}^{-2}$时，半导体中的空间电荷起伏有足够的时间和空间增长，形成向阳极传播的偶极层(畴)。偶极畴通常在阴极接触附近形成，因为那里掺杂起伏和电场扰动最大。成熟的偶极畴周期形成并在阳极消失，引起早期实验上观察到的耿(Gunn)振荡，转移电子器件也曾被称为耿效应器件，为人们所熟知。我们在前面已经详细讨论了这种模式。图 6.20(b)对于 n_0L 乘积为 $2.1\times10^{12}\text{cm}^{-2}$的器件表示了峰值电压时的电场分布，以及电压和电流的波形。如前所述，通过器件的渡越时间是 L/v_D，相应的振荡频率是 v_D/L。这种模式的最高转换效率对 GaAs 约为 10%，对 InP 约为 15%。

猝灭畴模式 在渡越时间模式中，器件上的电压主要降落在偶极畴上。当偏压降低时，畴的宽度也下降，甚至收缩为零，这发生于偏压 V_s。这样，当加在器件上的偏压降到 V_s 以下时，偶极畴猝灭；当偏压摆动回到高于阈值时，新的畴成核，过程重复。所以，如果偶极畴到达阳极之前猝灭，则谐振电路中 TED 的工作频率就可以高于渡越时间频率。图 6.20(c)对于一个以猝灭畴模式工作的器件表示了多畴形成以及电压和电流的波形(形成多个偶极畴是因为一个畴没有足够时间调节并吸收其他畴的电压)；该器件的 n_0L 乘积是 $4.2\times10^{12}\text{cm}^{-2}$，振荡频率大约是渡越时间频率的四倍。猝灭畴模式的效率对 GaAs 可达 13%，对 InP 可达 20%。

第七章　半导体激光器和光电二极管

半导体激光器(或激光二极管,简写为 LD)与其他激光器不同之处是它的体积小(约0.1mm长),而且能简单地用调制偏置电流的方法实现高频调制(因而能直接方便地载入高速率信息),是光纤通信最重要的光源。它也广泛用于光信息存储及高速激光打印等方面。此外,半导体激光器可以采用平面工艺加工,因而能与其他光学器件及电子学器件单片集成,构成复杂电路,即光电子集成电路(OEIC),这一技术将使光学线路和设备的性能大大提高,成本大大降低。本章主要叙述半导体激光器的基本原理和性能。光电二极管是在光通信等领域中广泛使用的一种光探测器件,在本章的最后一节讨论;该节还附带叙述了太阳电池的基本原理。

7.1　半导体激光器的基本结构和工作原理

7.1.1　基本结构(光学谐振腔)

最简单的半导体激光器由禁带较宽的 p 型和 n 型材料中间夹一层很薄的禁带较窄的另一种半导体材料(作为有源区)构成。激光器必须包含一个光学谐振腔,用以实现激光振荡所需的正反馈。对于半导体激光器,用得最广泛的是法布里-珀罗谐振腔(F-P 腔)。F-P 腔的主要特点是两端为彼此平行的抛光镜面(自然解理面),如图 7.1(a)所示。在自然解理面上,由于有源区的折射率 $\bar{n}$ 不同于空气的折射率 1 而构成反射镜。光在谐振腔中来回反射,只有波长 $\lambda/\bar{n}$(激光介质中的波长)满足下述条件的模式才能存在:

$$L = \frac{m}{2}\left(\frac{\lambda}{\bar{n}}\right), \quad m = 1,2,3,\cdots \tag{7.1}$$

或

$$\lambda = \frac{2\bar{n}L}{m} \tag{7.2}$$

L 是谐振腔的长度,λ 是真空中的光波波长。波长满足式(7.2)的谐振模式称为谐振腔的纵模。图 7.1(b)画出了 $m=1,2,3$ 的 3 个纵模。对式(7.2)微分可求得 F-P 腔纵模的间隔:

$$\Delta\lambda = \frac{\lambda^2}{2\pi L\left(1 - \frac{\lambda}{\bar{n}}\frac{\mathrm{d}\bar{n}}{\mathrm{d}\lambda}\right)} \tag{7.3}$$

如果忽略分母中的色散项(即令 $\mathrm{d}\bar{n}/\mathrm{d}\lambda = 0$),则有

$$\Delta\lambda = \lambda^2/2\bar{n}L \tag{7.4}$$

式(7.4)说明,腔长愈短,纵模间隔愈大。一般半导体激光器的腔长只有数百微米,比气体激光器和其他固体激光器要短得多,所以它的纵模间隔要宽得多。但是式(7.2)所表示的只是谐振腔所允许存在的纵模,它是一个无穷的系列,究竟激光器出现哪些纵模,还要看激光介质(有源区)的增益谱。只有那些增益达到阈值条件,而又被谐振腔允许的波长(或频率)才有激光振荡。由于半导体的增益谱较宽,谐振腔中常常有多个纵模起振。如果对激光器结构进行一些特殊的设计,可以只保留一个纵模,从而激光器发出较好的单色光,相干长度变得很大。F-P

腔的两个侧面是粗糙的，所以向两侧发出的光子不能反馈，谐振模式只有纵模。

图 7.1(c)表示由有源区和无源区(覆盖区，即 p 区和 n 区)组成的 3 层介质波导中介电常数 $\varepsilon(z)$的变化，这种变化使光波在 z 方向的传播受到限制。通常引入光限制因子 Γ 来描述有源区的光限制性质，它定义为有源区内的光能与有源区及无源区内总光能的比值：

$$\Gamma = \frac{\int_{\text{有源区}} |\mathscr{E}(z)|^2 \mathrm{d}z}{\int |\mathscr{E}(z)|^2 \mathrm{d}z} \tag{7.5}$$

$\mathscr{E}(z)$代表光波的电场。

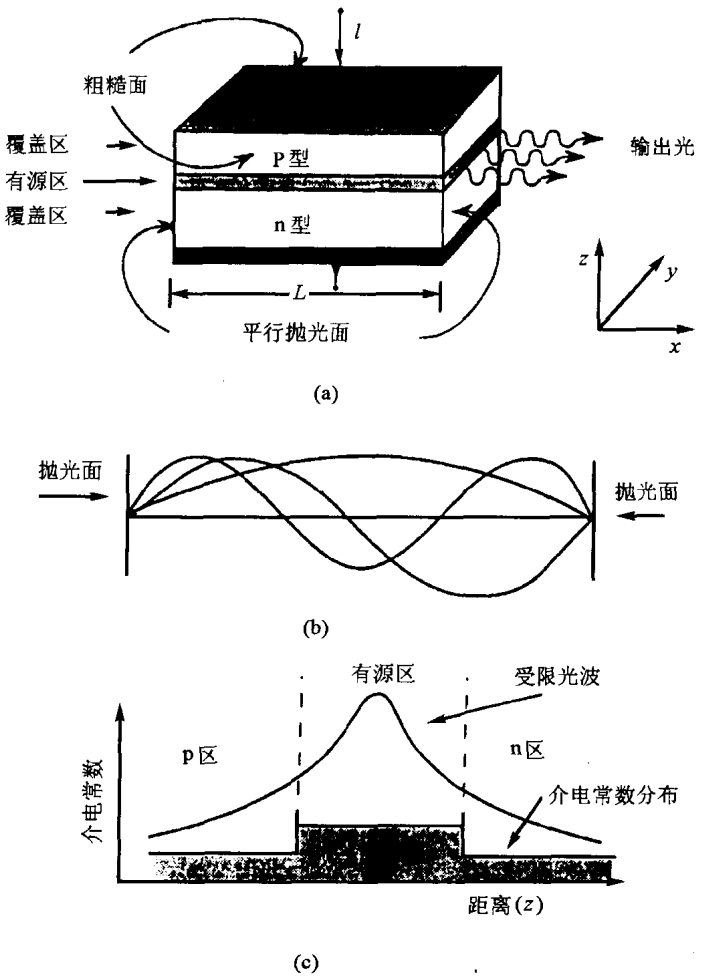


图 7.1 (a) 半导体激光器的基本结构(F-P 腔)；(b) 谐振模式；(c) 折射率变化

7.1.2 增益、损耗和光振荡条件

1. 光增益

当制作半导体激光器的 p-n 结正偏时，电子和空穴注入其间的有源区。如果注入的载流子浓度满足反转条件：

$$f_n(E_n) + f_p(E_p) > 1 \tag{7.6}$$

有源区成为增益介质。参与光跃迁过程的导带能态 E_n 及价带能态 E_p 与光子能量 $\hbar\omega$ 的关系为

$$\begin{aligned} E_n &= E_c + \frac{m_r}{m_n}(\hbar\omega - E_g) \\ E_p &= E_v - \frac{m_r}{m_p}(\hbar\omega - E_g) \end{aligned} \tag{7.7}$$

m_n 和 m_p 分别为电子和空穴的有效质量，m_r 为它们的折合质量。

光波在增益介质中行进时，光强（等于光子流密度乘以光子能量 $\hbar\omega$）和距离 x 的一般关系为

$$I_{ph} = I_{ph}^0 e^{G(\hbar\omega)x} \tag{7.8}$$

$G(\hbar\omega)$ 称为增益系数，根据第一章叙述的半导体光学性质，它可表示为

$$G(\hbar\omega) = \frac{\pi q^2 \hbar |p_{cv}|^2}{\bar{n} c m_0^2 \hbar\omega \varepsilon_0} g_{cv}(\hbar\omega)[f_n(E_n) + f_p(E_p) - 1] \tag{7.9}$$

或者，利用式(1.194)定义的自发寿命 $\tau_0\left(\tau_0^{-1} = \frac{\bar{n} q^2 \hbar\omega |p_{cv}|^2}{\pi m_0^2 \varepsilon_0 c^3 \hbar^2}\right)$，上式可改写为

$$G(\hbar\omega) = \frac{\hbar\lambda^2}{4\bar{n}^2 \tau_0} g_{cv}(\hbar\omega)[f_n(E_n) + f_p(E_p) - 1] \tag{7.10}$$

式中，$\lambda = 2\pi c/\omega$ 为真空中光波波长，$g_{cv}(\hbar\omega)$ 为联合态密度，

$$g_{cv}(\hbar\omega) = \frac{\sqrt{2} m_r^{3/2}}{\pi^2 \hbar^3}(\hbar\omega - E_g)^{1/2} \tag{7.11}$$

需要注意，对于半导体的情形，τ_0 应当用电子和空穴的辐射复合寿命 τ_r 取代。

增益系数 $G(\hbar\omega)$ 是注入载流子浓度 $n(=p)$ 的函数。要计算 $G(\hbar\omega)$，必须求出电子和空穴的准费米能级 E_{Fn} 和 E_{Fp}，以及它们占据能态的几率 $f_n(E_n)$ 和 $f_p(E_p)$。根据第 1.4 节，占据几率为

$$\begin{aligned} f_n(E_n) &= [1 + e^{(E_n - E_{Fn})/k_B T}]^{-1} \\ f_p(E_p) &= [1 + e^{-(E_p - E_{Fp})/k_B T}]^{-1} \end{aligned} \tag{7.12}$$

必须注意，激光器是在大注入（f_n 和 f_p 都较大）的条件下工作，占据几率不能由玻耳兹曼统计准确给出。对于给定的注入浓度 $n(=p)$，费米能级的位置可用 Joyce-Dixon 近似，由下式给出：

$$\begin{aligned} E_{Fn} &= E_c + k_B T[\ln\frac{n}{N_c} + \frac{1}{\sqrt{8}}\frac{n}{N_c}] \\ E_{Fp} &= E_v - k_B T[\ln\frac{p}{N_v} + \frac{1}{\sqrt{8}}\frac{p}{N_v}] \end{aligned} \tag{7.13}$$

N_c 和 N_v 分别为导带和价带的有效态密度，与温度 T 有关。

用上述有关的表达式，对于不同的注入载流子浓度 n，可以作为光子能量 $\hbar\omega$ 的函数来计算增益系数 $G(\hbar\omega)$。$G \sim \hbar\omega$ 曲线（增益谱）的一般形状如图 7.2 所示。由图可见，注入载流子浓度 $n < 1.5\times10^{18}\text{cm}^{-3}$ 时，$G(\hbar\omega)$ 是负的，表明 f_n 和 f_p 太小，尚未实现粒子数反转分布；但当 n 超过此值时，$G(\hbar\omega)$ 在一定频率范围内是正的，其峰值 G_{max} 随 n 增大并向短波方向移

动。G_{max}和 n 的关系可近似为

$$G_{max} \simeq \sigma_G(n - n_T) \tag{7.14}$$

式中 n_T 称为透明载流子浓度，σ_G 称为微分增益系数。对GaAs激光器，300K 时有 $n_T \simeq 1.55\times10^{18}cm^{-3}$和 $\sigma_G \simeq 1.5\times10^{-16}cm^2$。当 $\hbar\omega \gg E_g$ 时，由于与跃迁过程有关能态的占据几率很小，即使在十分高的注入载流子浓度下，$G(\hbar\omega)$也是负的。

微分增益系数 σ_G 随温度下降而增大，对 GaAs 激光器，300K 时为 $1.5\times10^{-16}cm^2$，而 77K 时为 $5\times10^{-16}cm^2$；原因在于 f_n+f_p 随注入载流子浓度 n 变化的速率随温度下降而增大。σ_G 还和半导体激光器的结构有关，量子阱激光器的 σ_G 比普通激光器的约大一倍；原因在于，量子阱的能态密度和体材料相比少了一个维度的贡献，较低的能态密度使得 f_n+f_p 随注入载流子浓度 n 变化的速率增大。对于激光器性能的讨论将会看到，高 σ_G 的激光器具有响应速度快和线宽小的优点。

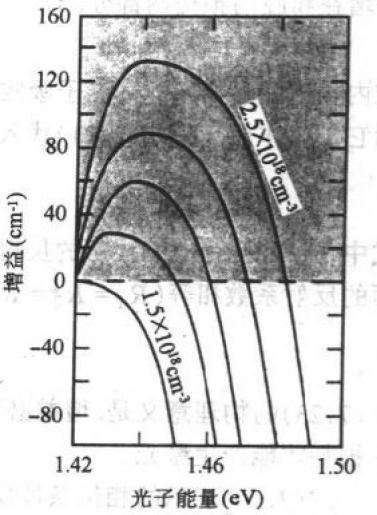


图 7.2 GaAs 激光器在不同注入载流子浓度下的增益系数

2. *光损耗和振荡条件*

光在沿谐振腔的传播中，一方面谐振腔内的光增益作用使光强增加，另一方面谐振腔内的损耗和端面损耗使光强减弱。为了获得激光，谐振腔内产生的光增益必须克服来自上述两方面原因的光损耗。所以，增益存在一个阈值，在阈值以下激光器不能工作，我们以 G_{th}表示阈值增益系数。还要注意，只有在有源区中传播的那部分光才能获得增益，即谐振腔内光的增益系数应修正为 $\Gamma G(\hbar\omega)$，Γ 为光限制因子。

可以通过考察一个平面电磁波在谐振腔内往返一次振幅和相位的变化导出实现激光的阈值条件。设频率 ω 的光波沿 x 方向(腔长方向)传播，则光的电场强度可表示为

$$\mathscr{E}(t) = \mathscr{E}_0 e^{i(\omega t - \tilde{n}k_0 x)} \tag{7.15}$$

式中，$k_0 = \omega/c = 2\pi/\lambda$，$c$ 和 λ 分别为真空中的光速和光波波长；$\tilde{n}$ 为有源区(激光介质)的复折射率，可表示为

$$\tilde{n} = \bar{n} - i\bar{k} \tag{7.16}$$

$\bar{n}$ 就是通常的折射率，只对相位改变起作用；$\bar{k}$ 则使传播衰减(或加强)，所以光的吸收(或增大)与 $\bar{k}$ 有关。包括腔内损耗在内的净增益系数 g 可表示为

$$g = -\frac{2\omega}{c}\bar{k} \quad 或 \quad g = -\frac{4\pi}{\lambda}\bar{k} \tag{7.17}$$

在长度为 L 的腔内往返一次，光场的变化为

$$r_1 r_2 e^{(g/2)\cdot 2L} \cdot e^{-i(2\pi\bar{n}/\lambda)\cdot 2L} \tag{7.18}$$

稳定条件下，光场的振幅(或光强)和相位都不改变，故有

$$r_1 r_2 e^{(g/2)2L} = 1 \quad (振幅条件) \tag{7.19}$$

和

$$e^{-i(2\pi\bar{n}/\lambda)2L} = 1 \quad (相位条件) \tag{7.20}$$

由式(7.19)表示的振幅条件可以得到激光振荡的阈值条件。在式(7.19)中，净增益系数 g 包

括增益和腔内损耗两部分，即

$$g = \Gamma G - \alpha_m \tag{7.21}$$

腔内损耗主要是光的载流子吸收以及缺陷引起的散射，在上式中我们用了吸收系数 α_m 来概括它们的作用。将式(7.21)代入式(7.19)，并以 G_{th}表示阈值增益系数，则激光的阈值条件为

$$\Gamma G_{th} = \alpha_m - \frac{1}{L}\ln(r_1 r_2) \tag{7.22}$$

式中 r_1 和 r_2 为电场强度的反射率，换成一般的反射率(光强的反射系数)R_1 和 R_2，并令两端面的反射系数相等($R_1 = R_2 = R$)，有

$$\Gamma G_{th} = \alpha_m - \frac{1}{L}\ln R \tag{7.23}$$

式(7.23)的物理意义是，增益必须克服介质的内部损耗和从谐振腔两个端面发射出去所造成的损耗才能产生激光。

式(7.20)表示的相位条件要求

$$\frac{4\pi\bar{n}}{\lambda}L = 2m\pi, \quad m = 1,2,3,\cdots \tag{7.24}$$

或

$$\lambda = \frac{2\bar{n}L}{m} \tag{7.25}$$

这一条件是对激光波长的限制，和前面从直观判断得出的式(7.2)一致。

7.2 半导体激光器的工作特性

7.2.1 阈值电流密度

设注入电流密度为 J，则电子(空穴)进入有源区的速率为

$$JA/q$$

注入的电子-空穴对的复合率为

$$nAd/\tau_r(J)$$

A 是谐振腔的面积，d 是有源区的厚度，$\tau_r(J)$是与电流密度有关的辐射复合寿命。在辐射复合效率为 1 的假设下，上述二者应当相等，从而得到

$$n = \frac{J\tau_r(J)}{qd} \tag{7.26}$$

阈值时，有

$$n_{th} = \frac{J_{th}\tau_r(J_{th})}{qd} \tag{7.27}$$

J_{th}表示阈值电流密度。正如在第一章讨论过剩载流子的辐射寿命时所指出的，τ_r 强烈依赖于载流子浓度(即注入电流密度)。随着注入电流密度超过 J_{th}，τ_r 之值变小。结果，虽然注入电流密度增加，有源区中的载流子浓度饱和且接近 n_{th}。方程(7.27)给出阈值电流密度 J_{th}和有源区厚度 d 之间的重要关系，为了得到低的阈值电流密度，有源区厚度应当减小。但从有源区内每秒产生的光子数(等于 $n_{th}Ad/\tau_r$)来看，为了得到大的激光输出，有源区厚度应当很大。阈值电流密度指器件产生激光输出的最小注入电流密度，与激光器材料、工艺、结构等因素有关，还随温度升高而增大。降低阈值电流密度一直是半导体激光器工艺改进的方向之一。

J_{th}和阈值增益 G_{th}相对应，估算 J_{th}时可以从下式出发：

$$G_{th} = \sigma_G(n_{th} - n_T) \tag{7.28}$$

并利用式(7.23)和(7.27)，将阈值电流密度表示(设光限制因子 $\Gamma=1$)为

$$J_{th} = q\left[n_T + \frac{\alpha_m - \frac{1}{L}\ln R}{\sigma_G}\right]d/\tau_r \tag{7.29}$$

设 $\alpha_m = 10\ \mathrm{cm}^{-1}$，$L = 500\ \mu\mathrm{m}$，有源区材料为 GaAs($n_T \simeq 1.55\times10^{18}\mathrm{cm}^{-3}$，$\sigma_G \simeq 1.5\times10^{-16}\mathrm{cm}^2$，$\tau_r \simeq 5\ \mathrm{ns}$)，则有

$$J_{th}/d \simeq 4\ 830\ \mathrm{A/cm}^2$$

双异质结激光器的有源区厚度(1000～3000 Å)比同质 p-n 结激光器的小得多，J_{th}降低了一二个数量级，使半导体激光器在室温下连续工作成为可能。量子阱激光器的阈值电流密度更是低得多，这不仅由于它的有源区十分薄(～100Å)，而且由于二维情形下的低能态密度使较少的注入载流子即可实现粒子数反转(即 n_T 较小)。

J_{th}和温度 T 的关系可表示为

$$J_{th}(T) = J_0 e^{T/T_0} \tag{7.30}$$

T，T_0 都以绝对温度表示，T_0 称为激光器的特征温度，J_0 为 $T = T_0$ 时阈值电流密度的 1/e。T_0 表征半导体激光器的阈值电流对温度的敏感性，T_0 越大，器件的温度稳定性越好。InGaAsP半导体激光器的 T_0 典型值在 50～70K，而 GaAs 激光器的 T_0 超过 120K。因此，长波长器件的温度稳定性不如短波长器件。对于 1.55 μm 的 InGaAsP 半导体激光器，温度超过 100℃时一般不能再产生激光输出。

7.2.2 输出功率和效率

图 7.3 表示半导体激光器输出功率 P_0 和注入电流强度 I 之间的典型关系。如前所述，只有注入电流 I 超过阈值(这里以 I_{th}表示)，才会有激光输出。$I < I_{th}$时，注入载流子的数量较小，谐振腔内增益不足以克服损耗，结果器件是自发发射出光。令 β 表示不能从器件逸出的光子所占比例，则光子流输出为

$$I_{ph} = (1-\beta)R_{sp}Ad = (1-\beta)I/q \tag{7.31}$$

R_{sp}是自发发射速率。由于光子损耗项 β 的数值很高，光子输出 I_{ph}很小。同时，自发发射的光谱线很宽，相干性很差。

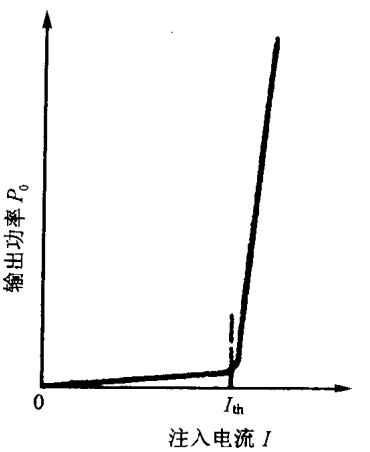


图 7.3 半导体激光器的输出功率 P_0 和注入电流 I 的关系

当 $I > I_{th}$时，激光器输出功率 P_0 随 I 急剧增加。令谐振腔有源区内注入的电子-空穴对发生辐射复合的几率为 η_i(η_i 亦即有源区内每秒产生的光子数与每秒注入的载流子数之比，这一比值称为内量子效率)，则可对受激发射的光功率写出下面的表达式：

$$P = \frac{(I - I_{th})\eta_i}{q}\hbar\omega \tag{7.32}$$

此功率的一部分耗散在激光器内部，其余部分通过两端的反射面透射出去。从阈值条件的表

达式(7.23)不难看出,这两部分功率分别正比于 α_m 和 $\frac{1}{L}\ln\frac{1}{R}$,故输出功率(即透射功率)为

$$P_0 = \frac{(I - I_{th})\eta_i \hbar\omega}{q} \cdot \frac{\frac{1}{L}\ln\frac{1}{R}}{\alpha_m + \frac{1}{L}\ln\frac{1}{R}} \tag{7.33}$$

定义微分外量子效率 η_{ex} 为光子输出速率的增量与注入速率(每秒注入的电子-空穴对数)的增量之比,即

$$\eta_{ex} = \frac{d(P_0/\hbar\omega)}{d[(I - I_{th})/q]} \tag{7.34}$$

将式(7.33)代入上式,可得

$$\eta_{ex}^{-1} = \eta_i^{-1}\left[\frac{\alpha_m L}{\ln\frac{1}{R}} + 1\right] \tag{7.35}$$

利用式(7.35)以及由实验测出的 η_{ex} 和 L 的关系可以确定内量子效率 η_i,G_aA_s 的 η_i 大约是 0.9～1.0。

若加到激光器的正向电压为 V,输入电功率就是 VI,则激光器将电功率转换成光输出功率 P_0 的效率为

$$\eta = \frac{P_0}{VI} = \eta_i\left(\frac{I - I_{th}}{q}\right)\left(\frac{\hbar\omega}{qV}\right)\frac{\ln\frac{1}{R_1}}{\left(\alpha_m + \ln\frac{1}{R}\right)} \tag{7.36}$$

理论上,$qV \simeq \hbar\omega$。实际上,由于激光二极管串联电阻 R_s(包括材料的体电阻和电极接触的电阻),提供给每一个注入载流子的能量 qV 大于光子的能量 $\hbar\omega$,使转换效率 η 降低。一般,$qV \sim 1.4E_g$,$\hbar\omega \simeq E_g$,已经做到 300 K 时 $\eta \sim 30\%$。

7.2.3 调制特性

半导体激光器的光输出必须受到调制才能用于信息传输。最简单也是最重要的调制方法是由注入电流直接调制,这对于实现激光器和调制电路的单片集成十分重要。依赖于应用,可以把调制技术划分为下述三种方式。

(1) 大信号调制　在这种方式中,注入电流的变化范围从低于阈值到远高于阈值,从而实现激光器关断和开启之间的转换。这类调制可用于"光互连"或逻辑应用。以后将看到激光器对这种调制的响应很慢(～10 ns)。大信号调制由于响应慢和输出谱宽,不能用于光通信。

(2) 小信号调制　在这种方式中,激光器偏置电流远高于阈值,而调制电流的幅度很小。对于特殊设计的高速半导体激光器,小信号调制的带宽可以达到 50 GHz,但实际上调制频率通常小于 10 GHz,主要受限于电子线路而不是激光器芯片本身。

(3) 脉冲码调制　这种调制广泛用于现代光通信系统,对它的分析介乎大、小信号调制之间。在这种调制中,激光器偏置远高于阈值,再以较大的矩形脉冲调制,但注入电流始终高于阈值。用脉冲码调制方式,激光器可以达到 10 Gb/s 的速率。

1. 大信号瞬态响应

激光器的大信号开关涉及二极管电流从低于阈值到高于阈值,图 7.4(a)给出了关于瞬态

响应的一个计算结果。在电流脉冲之前，激光器有源区内的载流子浓度实际上是零。随着电流脉冲接通，载流子浓度增加，器件内的光增益开始增加，但在达到损耗值之前很少有光子逸出谐振腔。这样，必须经过延迟时间 t_d 才会有光子从器件出来。另一方面，一旦载流子浓度达到 n_{th}，受激发射开始；但在稳态之前，当载流子浓度值超过 n_{th} 时，光子输出将超过稳定值。高额的光子输出反过来降低载流子浓度，因为这时有较多的电子-空穴对复合。这样，载流子浓度和光子输出产生振荡（弛豫振荡）。

我们通过载流子浓度 n 变化的速率方程来计算延迟时间 t_d，这一方程可写作

$$\frac{dn}{dt} = \frac{J}{qd} - \frac{n}{\tau} - R_{st} \tag{7.37}$$

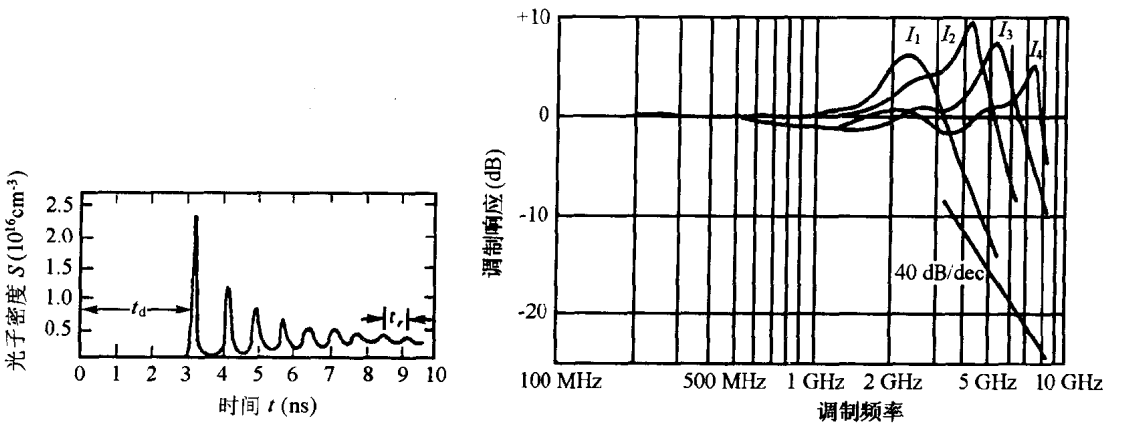
$\tau = \tau_r \tau_{nr}/(\tau_r + \tau_{nr})$ 是电子-空穴对的复合时间，其中 τ_r 是辐射复合寿命，τ_{nr} 是非辐射复合寿命。上式右方的第一项表示载流子流入有源区的速率，第二项表示由于自发发射载流子的损失速率，第三项表示由于受激发射的损失速率。如果电流密度从 0 到 J 变化的时间内 $n < n_{th}$，则谐振腔内没有光子存在，$R_{st} \sim 0$。从 $t=0$ 到 t 和 $n(0)$ 到 $n(t)$ 积分这一方程，得到

$$t = \tau \ln\left(\frac{J - qn(0)d/\tau}{J - qn(t)d/\tau}\right) \tag{7.38}$$

当 $n(t) = n_{th}$ 时，光子密度开始变化。这样，延迟时间（设 $n(0)=0$）为

$$t_d = \tau \ln\left(\frac{J}{J - qn_{th}d/\tau}\right) = \tau \ln\left(\frac{J}{J - J_{th}}\right) \tag{7.39}$$

如果非辐射过程可以忽略，则 $\tau = \tau_r$，这样在电流脉冲接通和光子从激光器出现之间有几个纳秒的延迟时间，这一时间以及光子开始出现以后产生的弛豫振荡对于激光器的许多应用是一个严重障碍。



(a) 大信号开关的瞬态响应　　(b) 不同偏置电流 ($I_4 > I_3 > I_2 > I_1$) 下归一化调制光输出与调制频率的关系

图 7.4　半导体激光二极管的调制响应

2. 小信号调制

在小信号调制中，激光器偏置通常远高于阈值，以得到最大有效带宽（以后将看到这一点）。在如此高的偏置下，激光器主模占优势。所以，在下面的讨论中，我们假定只存在单一的激光模式；同时假定载流子均匀注入，并忽略光子密度沿谐振方向的起伏。在这些条件下，有源区内光子密度 S 和注入电子（空穴）浓度变化的速率方程为

$$\frac{\mathrm{d}S}{\mathrm{d}t} = \left[\frac{\Gamma\sigma_G(n - n_T)}{1 + \varepsilon S} - \alpha_c\right]Sv_g + \Gamma\frac{\beta n}{\tau_r} \tag{7.40}$$

$$\frac{\mathrm{d}n}{\mathrm{d}t} = \frac{I}{qV} - \frac{n}{\tau_r} - \frac{\sigma_G(n - n_T)}{1 + \varepsilon S}Sv_g \tag{7.41}$$

在以上两个方程中，$v_g = \mathrm{d}\omega/\mathrm{d}k$ 是光的群速度，与折射率 $\bar{n}$ 的关系可表示为 $v_g = \frac{c}{\bar{n}}\left(1 + \frac{\lambda}{\bar{n}}\frac{\mathrm{d}\bar{n}}{\mathrm{d}\lambda}\right)$，忽略介质的色散效应时有 $v_g = c/\bar{n}$；Γ 是光限制因子，表示有源区内光子数在总光子数中所占比例；τ_r 是载流子的自发发射寿命；$\alpha_c = \alpha_m - \frac{1}{L}\ln R$，是谐振腔的光子损耗率，包括腔内损耗和端面损耗；$\sigma_G(n - n_T)/(1 + \varepsilon S)$是考虑了增益饱和效应以后的增益系数，式中 ε 是表征增益同光子密度非线性关系的参量(增益饱和效应是粒子数反转分布随光子密度改变而导致增益下降的效应)；β 是自发发射因子，表示自发发射光子总数中进入单一激光模式的比率，由于自发发射的能量范围是 k_BT 的量级，而激光线宽只是几个 μeV，β 的典型值为 $10^{-4}\sim10^{-5}$；I 是注入电流强度，这里应理解为注入电流的辐射部分(一个重要的非辐射复合过程是俄歇过程。在俄歇过程中，电子-空穴对复合时把它们的能量交给第三个电子或空穴，结果没有光子发射。俄歇过程在窄带隙激光器中特别重要，其重要性随温度增加。)；V 是有源区的体积。方程(7.40)的物理意义是，有源区光子密度增加的速率等于因受激发射和自发发射增加的速率减去因腔损耗而减少的速率，方程(7.41)的物理意义类似。

为了避免数学上的繁琐而不影响对半导体激光器基本调制特性的讨论，下面的分析中将忽略增益饱和效应(即令 $\varepsilon=0$)和自发辐射对激光光子密度的影响(令 $\beta=0$)。于是，速率方程为

$$\frac{\mathrm{d}S}{\mathrm{d}t} = \Gamma A(n - n_T)S - \frac{S}{\tau_{ph}} \tag{7.42}$$

$$\frac{\mathrm{d}n}{\mathrm{d}t} = \frac{I}{qV} - \frac{n}{\tau_r} - A(n - n_T)S \tag{7.43}$$

式中，$A = \sigma_G v_g$，τ_{ph}为谐振腔内的光子寿命。因为谐振腔光子在单位时间内损失的几率为 $\alpha_c v_g$，光子的平均寿命可表示为

$$\tau_{ph} = \frac{1}{\alpha_c v_g} \simeq \frac{\bar{n}}{\alpha_c c} \tag{7.44}$$

c 是真空中的光速，$\bar{n}$ 是有源区的折射率。

令式(7.42)和(7.43)的左边等于零，即得到速率方程的稳态解：

$$0 = \Gamma A(n_o - n_T)S_0 - S_0/\tau_{ph} \tag{7.45}$$

$$0 = \frac{I}{qV} - \frac{n_o}{\tau_r} - \Gamma A(n_0 - n_T)S_0 \tag{7.46}$$

半导体激光器的调制响应一般只能对速率方程进行数值求解得到，但对交流小信号调制可以得到解析结果。交流小信号调制是信号调制电流正弦变化而激光器的偏置电流 I_0 超过阈值 I_{th}，并且调制电流的幅度 i_1 满足 $i_1 \ll I_0$，即注入电流为

$$I = I_0 + i_1 e^{i\omega t} \tag{7.47}$$

在小信号调制下，载流子浓度 n 和光子密度 S 可表示为

$$n = n_0 + n_1 e^{i\omega t}, \quad S = S_0 + S_1 e^{i\omega t} \tag{7.48}$$

其中 S_0 和 n_0 满足方程(7.45)和(7.46)。

将式(7.48)代入方程(7.42)和(7.43)，忽略二阶小量，并利用方程(7.45)和(7.46)，得到下面两个方程：

$$-i\omega n_1 = \frac{-i_1}{qV} + \left(\frac{1}{\tau_r} + AS_0\right) n_1 + \frac{S_1}{\Gamma\tau_{ph}} \tag{7.49}$$

$$i\omega S_1 = \Gamma A S_0 n_1 \tag{7.50}$$

我们主要感兴趣的是调制响应 $S_1(\omega)/i_1(\omega)$，从以上二式有

$$S_1 = \frac{(i_1/qV)\Gamma A S_0}{\omega_0^2 - \omega^2 + i\omega\gamma_d} \tag{7.51}$$

从上式可得到调制响应函数为

$$H(\omega) \equiv \frac{S_1(\omega)}{S_1(0)} = \frac{\omega_0^2}{\omega_0^2 - \omega^2 + i\omega\gamma_d} \tag{7.52}$$

其中

$$\omega_0 = \sqrt{\frac{AS_0}{\tau_{ph}}} \tag{7.53}$$

$$\gamma_d = \frac{1}{\tau_r} + AS_0 \tag{7.54}$$

$H(\omega)$在下述条件下出现峰值：

$$\omega = i\frac{\gamma_d}{2} \pm \sqrt{\omega_0^2 - \frac{\gamma_d^2}{4}} \tag{7.55}$$

方程(7.55)中的第一项为阻尼项，第二项是响应峰值所对应的频率，称为弛豫共振频率，以 ω_R 表示，为

$$\omega_R = \sqrt{\frac{AS_0}{\tau_{ph}} - \frac{1}{4}\left(\frac{1}{\tau_r} + AS_0\right)^2} \tag{7.56}$$

$\tau_r \sim 4\times10^{-9}$ s，同时可以证明：$\tau_{ph} \sim 10^{-12}$ s，$AS_0 \sim 10^{-9}$ s，所以下式是一个很好的近似：

$$\omega_R = \sqrt{\frac{AS_0}{\tau_{ph}}} = \sqrt{\frac{\sigma_G}{\tau_{ph}}\left(\frac{I - I_{th}}{qV}\right)\eta_i} \tag{7.57}$$

上式的最后一步能够从有源区产生的光功率既可表示为 $S_0\hbar\omega v_g V$ 也可表示成$[(I - I_{th})/q]\hbar\omega\eta_i$得到证明。

由方程(7.52)可以看出，$\omega < \omega_R$ 时 $H(\omega) \sim 1$，响应曲线基本上是平坦的；直到 $\omega = \omega_R$ 时出现峰值，其大小为 ω_R/γ_d；然后随调制频率 ω 增加而迅速下降。所以 ω_R 是半导体激光器的最高响应频率(响应带宽)。图 7.4(b)给出了激光器典型的调制响应曲线。

下面简短讨论影响半导体激光器频率响应的因素。

(1) 注入电流和阈值电流

从式(7.57)十分清楚，激光器的响应带宽 ω_R 随光子密度 S_0 增大；或者说，若激光器偏置在大注入电流 I 下，器件将工作得更好。由于相同注入电流下，低阈值激光器的光子密度大，低阈值电流激光器是很重要的。

但是，不能在十分大的电流下简单地驱动激光器。在大注入下，高温和高光子密度引起的效应(例如增益饱和效应)将使激光器性能退化。

(2) 光子寿命

响应带宽 ω_R 的表达式(7.57)表明,光子寿命 τ_{ph}应尽可能短。光子寿命可以用较短的谐振腔来降低,但这会增加损耗,从而必须以较大的 n_{th}来达到相同的增益。这样,对于一定的激光器,有一个最佳腔长,典型值约为 100 μm。

(3) 微分增益

微分增益 σ_G 在式(7.57)中的出现表明,响应带宽可以通过选择激光器的有源区(例如量子阱激光器)使之有高微分增益而得到改善。

(4) 俄歇效应

俄歇效应对激光器性能有两点重要影响。一方面,由于一部分电流不能用来产生光子,必须以大电流来驱动激光器来达到一定的 S_0 值;另一方面,由于载流子寿命 τ 减小,阻尼系数 γ_d 增加,器件响应受到损害。

7.2.4 激光线宽

同其他光源一样,半导体激光器产生的单一模式的光并不是单一波长(或频率)的光,而是具有一定的波长范围,即具有一定的光谱线宽度。激光线宽越小,它的相干长度就越大,这对长距离的光通信系统和高分辨率的探测系统等都是十分重要的。所以激光线宽是表征激光器性能的重要参数之一。

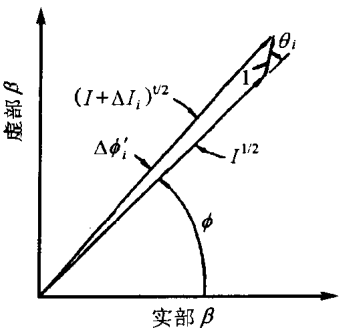


图 7.5 由第 i 个自发发射事件引起的光场位相 ϕ 和光场强度 I 的瞬时变化

激光线的宽度主要来自激光场的位相起伏,而这种起伏则来自自发发射所引起的激光场的位相和强度无规则变化。如图 7.5 所示。每一次自发发射引起的场强度化,平均来说等价于使光场增加一个光子。为了恢复稳态光强,激光器将经历一个弛豫过程(持续时间约为 1 ns)。在此期间,将有增益变化 $\Delta G(t)=(-2\omega/c)\Delta\bar{k}(t)$,其中 $\Delta\bar{k}$为折射率的虚部相对于稳态值的微分增量。$\bar{n}$(折射率实部)和 $\bar{k}$ 的变化由载流子浓度 n 的变化引起,二者变化之比为

$$\alpha = \Delta\bar{n}/\Delta\bar{k} \tag{7.58}$$

$\Delta\bar{n}$将导致激光场的附加相移和激光线展宽。利用 $\Delta G=\left(-\frac{4\pi}{\lambda}\right)\Delta\bar{k}$可将式(7.58)写作

$$\alpha = -\left(\frac{4\pi}{\lambda}\right)\left(\frac{\mathrm{d}\bar{n}/\mathrm{d}n}{\mathrm{d}G/\mathrm{d}n}\right) \tag{7.59}$$

α 称为线宽增益因子,其中 $\mathrm{d}G/\mathrm{d}n=\sigma_G$ 为微分增益系数。

在图 7.5 中,光场用复振幅 β 表示,谐振腔内平均光强 $I=\beta^*\beta$ 用平均光子数表示,$\phi(t)$则表示光场的位相。第 i 个自发发射事件使光场增加一个光子,故引起光场的改变为

$$\Delta\beta_i = \mathrm{e}^{i(\phi+\theta_i)} \tag{7.60}$$

其中 $\Delta\beta_i$ 的大小等于 1,而具有无规则的位相$(\phi+\theta_i)$,因为 θ_i 是无规则的。

单个自发发射使激光场位相起伏 $\Delta\phi_i$ 来自两个方面的贡献,它们分别用 $\Delta\phi_i'$ 和 $\Delta\phi_i''$ 表示。$\Delta\phi_i'$来自 $\Delta\beta_i$ 的异相分量,即与位相 ϕ 的变化方向一致的分量,这一分量(在 $I^{1/2}\gg1$ 的情形下)为

$$\Delta\phi_i' = I^{-1/2}\sin\theta_i \tag{7.61}$$

$\Delta\phi_i''$ 来自沿强度变化方向的分量,与光场的强度变化有关。从图 7.5 看到,光场振幅从 $I^{1/2}$变

到$(I+\Delta I_i)^{1/2}$。应用余弦定理得到

$$\Delta I_i = 1 + 2I^{1/2}\cos\theta_i \tag{7.62}$$

为了把 $\Delta\phi_i''$ 和 ΔI_i 联系起来，我们利用 I 和 ϕ 变化的速率方程

$$\frac{dI}{dt} = \left(Gv_g - \frac{1}{\tau_{ph}}\right)I \tag{7.63}$$

$$\frac{d\phi}{dt} = \frac{\alpha}{2}\left(Gv_g - \frac{1}{\tau_{ph}}\right) \tag{7.64}$$

其中 Gv_g 是受激发射的速率，$1/\tau_{ph}$表示谐振腔内部和端面光子损耗的速率。将方程(7.63)和(7.64)联立，有

$$\frac{d\phi}{dt} = \frac{\alpha}{2I}\frac{dI}{dt} \tag{7.65}$$

对方程(7.65)积分，并考虑 $I(0)=I+\Delta I_i$ 和 $I(\infty)=I$，得到

$$\Delta\phi_i' = \left(-\frac{\alpha}{2I}\right)\Delta I_i = \left(-\frac{\alpha}{2I}\right)\left[1 + 2I^{1/2}\cos\theta_i\right] \tag{7.66}$$

总的位相变化为

$$\Delta\phi_i = \Delta\phi_i' + \Delta\phi_i'' = -\frac{\alpha}{2I} + I^{-1/2}\left[\sin\theta_i - \alpha\cos\theta_i\right] \tag{7.67}$$

第一项是一个小的常数项，它引起来自自发发射的频率漂移。忽略常数项，对于 $N=R_{sp}t$ (R_{sp}是自发发射速率)个自发发射，总的位相起伏为

$$\Delta\phi = \sum_i I^{-1/2}\left[\sin\theta_i - \alpha\cos\theta_i\right] \tag{7.68}$$

从上式可计算 $\Delta\phi^2$ 之值。令交叉项趋于零(由于 θ_i 无规则，它在 0 到 2π 之间取值的几率相等)，得到

$$\overline{\Delta\phi^2} = \frac{R_{sp}t}{2I}(1+\alpha^2) \tag{7.69}$$

激光线强度按频率分布通常用洛仑兹函数描述，半功率点之间的宽度为

$$\Delta\nu = \frac{1}{\pi t_c} \tag{7.70}$$

t_c为每经受一次自发发射以后激光场恢复稳态强度的弛豫时间，由下式给出(可参阅文献[18]，p.377～385)：

$$\frac{1}{t_c} = \frac{\overline{\Delta\phi^2}}{2t} \tag{7.71}$$

结合式(7.69)，(7.70)和(7.71)，得到激光线宽度

$$\Delta\nu = \frac{R_{sp}}{4\pi I}(1+\alpha^2) \tag{7.72}$$

光强 I 可以用激光器的输出功率表示。对于两个端面输出功率相等的激光器，I 和一个端面输出功率P_0 的关系为

$$I = \frac{2P_0}{h\nu v_g\alpha_c'} \tag{7.73}$$

式中 v_g 是光的群速，α_c' 是端面损耗系数($\alpha_c' = \frac{1}{L}\ln\frac{1}{R}$)。

自发发射速率 R 和增益 G 的关系为

$$R_{sp} = n_{sp} G v_g \tag{7.74}$$

n_{sp}为进入单一激光模式中的自发发射速率与受激发射速率之比，可表示为

$$n_{sp} = \frac{f_n f_p}{f_n + f_p - 1} = [1 - e^{[h\nu - (E_{Fn} - E_{Fp})]/k_B T}]^{-1} \tag{7.75}$$

稳态时，有 $G = \alpha_m + \alpha_c'$，所以最终得到激光线宽为

$$\Delta\nu = v_g^2 \left(\frac{n_{sp} h\nu}{8\pi P_o} \right) \alpha_c' (\alpha_m + \alpha_c')(1 + \alpha^2) \tag{7.76}$$

要减小激光线的宽度，最主要的是要降低线宽增益因子 α，由式(7.59)可知，应当选用微分增益系数 σ_G 大的激光介质，例如选用量子阱激光器。

7.3 双异质结(DH)激光器

这是至今为止最有实用价值的半导体激光器，其中最为常见的是 AlGaAs/GaAs DH 激光器和 InP/InGaAsP DH 激光器。三元合金 AlGaAs 的禁带宽度比 GaAs 的大，被用作对注入 GaAs 有源区(激活区)中电子和空穴的限制层，激光波长和 GaAs 的禁带宽度相对应($\lambda \simeq 0.89\,\mu m$)；InGaAsP 是四元合金，其发射的激光波长覆盖光纤损耗最小的波长窗口(1.3～1.6 μm)。本节简要介绍这两种激光器的基本知识。

7.3.1 能带图和光场分布

以 AlGaAs/GaAs DH 激光器为例，图 7.6(b)表示出了 DH 激光器的能带图(加有正向偏压 V)、折射率变化和光强分布。为了比较，图 7.6(a)表示了同质结激光器相应的特性。

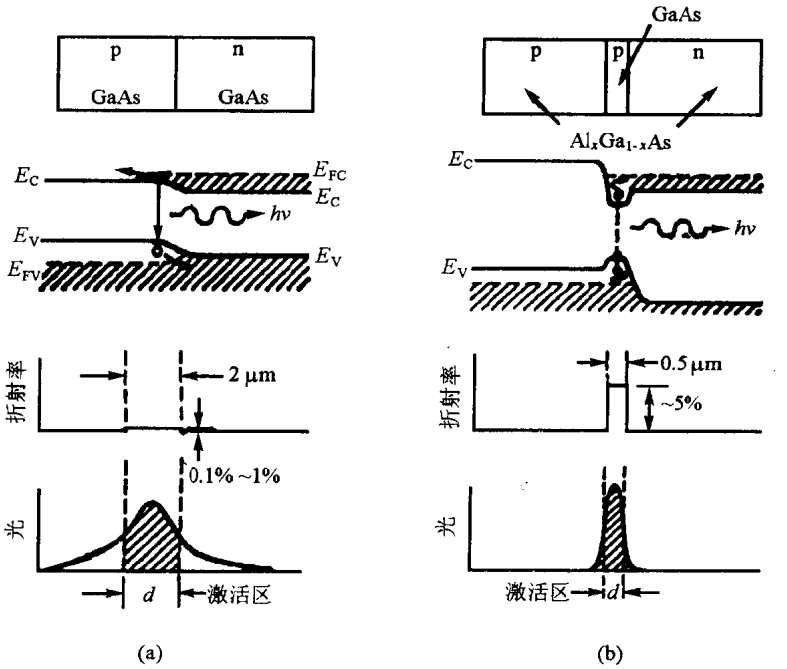


图 7.6 同质结激光器(a)和 DH 激光器(b)的能带图、折射率变化、光强分布

三元合金 AlGaAs 由 AlAs 和 GaAs 这两种Ⅲ-Ⅴ族化合物混合而成，实质上是 GaAs 中一部分 Ga 原子被 Al 原子取代。以 x 表示 Al 原子数与Ⅲ族(或Ⅴ族)总原子数之比，则这种三元合金表示为 $Al_xGa_{1-x}As$；其禁带宽度和折射率与组分比例 x 有关。GaAs 作为有源区做得很薄，其厚度小于 0.5 μm，通常是不掺杂的，以减少光子的损耗。$Al_xGa_{1-x}As$ 的禁带宽度 $(0<x<0.45)$可表示为

$$E_g = 1.424 + 1.247x \quad [\text{eV}] \tag{7.77}$$

它比 GaAs 的大，所以载流子被异质结能带边的不连续(ΔE_c 或 ΔE_v)所形成的势垒限制在 GaAs(有源区)内，从而有很高的辐射复合效率。而在同质结激光器内，载流子能够离开产生辐射复合的有源区，因此发光效率不高。另一方面，GaAs 的折射率比 $Al_xGa_{1-x}As$ 的大，其差值可表示为

$$\bar{n}_{GaAs} - \bar{n}_{Al_xGa_{1-x}As} \simeq 0.62x \tag{7.78}$$

所以两个限制层($p-Al_xGa_{1-x}As$ 和 $n-Al_xGa_{1-x}As$)和夹在它们中间的有源层(GaAs)将构成三层介质波导。由于有源层的折射率比限制层的大，当光线射向限制层的入射角大于临界角时将发生全反射，也就是说光线被导向平行于介质层界面的方向而大大地被约束在有源层内(光限制因子接近于 1)。而在同质结激光器内，有源层内外折射率之差 $\Delta\bar{n}$ 较小，光限制因子也就较小。

7.3.2 半导体材料

制作半导体激光器必须考虑的主要因素之一是选择能够发射所需波长的半导体材料。由于光子能量 $h\nu$ 近似等于半导体的禁带宽度 E_g，激光波长可以从 $E_g \simeq hc/\lambda$ 得到，其中 h 是普朗克常数，c 是真空中的光速。如果 E_g 用 eV 表示，λ 用 μm 表示，则

$$\lambda = \frac{1.24}{E_g} \tag{7.79}$$

对于限制层，应选择禁带宽度约大 200 meV 的半导体材料。在 AlGaAs/GaAs DH 激光器中，与 GaAs 的禁带宽度对应的激光波长近似为 0.89 μm。在 InP/InGaAsP DH 激光器中，$In_{1-x}Ga_xAs_yP_{1-y}$是四元合金，对于同 InP 晶格匹配($y=2.2x$)的情形，其禁带宽度可表示为

$$E_g = 1.35 - 0.72y + 0.12y^2[\text{eV}], \quad 0 \leqslant y \leqslant 1 \tag{7.80}$$

这一材料体系覆盖了从 0.92 μm 至 1.65 μm 的波长范围。光纤的最低损耗位于 1.3～1.6 μm 波长范围，所以 InP/InGaAsP DH 激光器在长距离光纤通信系统中是至关重要的。AlGaAs/GaAs DH 激光器常用于短距离(小于 2 公里)的光纤通信系统。

图 7.7 给出了几种二元、三元和四元半导体材料的禁带宽度和晶格常数，图中的黑点表示二元半导体，它们之间的连线对应它们构成的三元半导体材料，阴影部分则表示 Ga, In, As, P 构成的四元半导体材料。为了制成晶格匹配的异质结，只能采用图中基本上在同一竖直线(虚线)上的半导体材料，它们具有几乎相同的晶格常数，但禁带宽度不同。图中 GaAs 和 AlAs 的连线基本上是竖直线，线上的材料 $Al_xGa_{1-x}As$ 对任何摩尔分数 x 都能同 GaAs 实现晶格匹配，而当 $x<0.45$ 时都是直接禁带半导体，可以构成 $Al_xGa_{1-x}As/GaAs$ 异质结。而对于 $In_{1-x}Ga_xAs_yP_{1-y}$四元半导体，只有 y 和 x 的比值被限定为 $y/x=2.2$ 时，其晶格常数落在通过 InP 的竖直线上，从而能和 InP 构成晶格匹配的异质结。

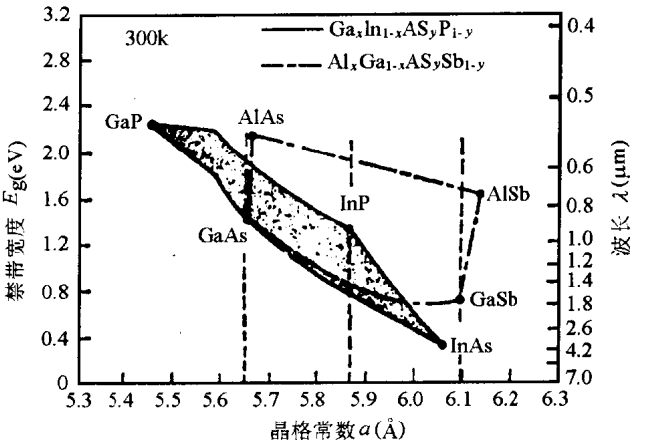


图 7.7　几种半导体材料的禁带宽度和晶格常数

如果异质结生长工艺不够合理，或者材料间的晶格匹配不很完善，则会在异质结界面处形成许多位错或其他缺陷，这些缺陷起着非辐射复合中心的作用，降低激光器的效率。在激光器的使用过程中，异质结界面处的缺陷和应力会使有源区内部增加一些不发光的暗区，使激光器的工作寿命缩短。所以，研究异质结生长工艺和晶格匹配，提高异质结界面的完整性，是研制长寿命激光器的关键。

7.3.3　改进的器件结构

1. 条形激光器

在前面叙述的 DH 激光器中，载流子和光子在垂直于结平面的方向（纵向）上受到限制（双异质结的性能），但在结平面内不受限制，整个结平面都能发射激光，这种激光器称为宽面激光器。为了降低工作电流和改善激光器输出特性，实际的半导体激光器通常做成条形结构，即在结平面（横向）上再对载流子和光子进行限制，这种激光器称为条形激光器。条形激光器的结构有两种基本类型，即增益导引和折射率导引，如图 7.8 所示。该图是半导体激光器在光输出方向的断面图（衬底采用 GaAs, InP 等材料，P, N 表示禁带宽度比有源层的要宽的 p, n 型材料）。

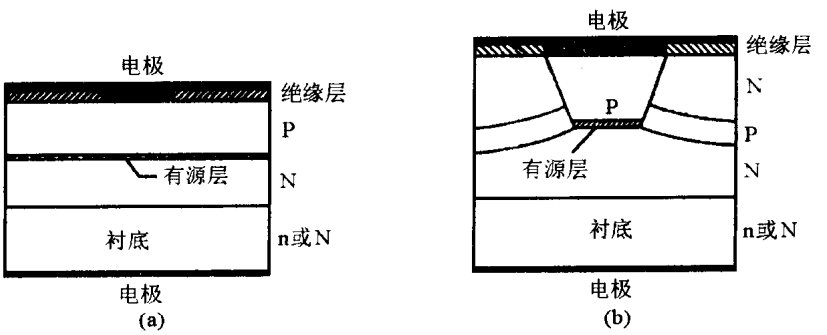


图 7.8　(a) 增益导引激光器；(b) 折射率导引激光器

增益导引半导体激光器通过在结平面方向（横向）上注入载流子浓度分布（增益分布）对光

子进行限制。在图 7.8(a)所示的结构中，除接触区之外，全部用氧化层绝缘，因此注入电流和发光区域都被限制在金属接触下面的狭窄条形区域(条区宽度的典型值 5-10 μm)内，不像宽面 DH 激光器那样在有源区的整个端面都发光。

折射率导引半导体激光器通过在横向引入一个折射率差 $\Delta\bar{n}$，形成与纵向的双异质结相类似的波导效应而实现对光子的限制。在图 7.8(b)所示的结构中，条形有源区(例如 p-GaAs 区)被折射率较低的宽禁带材料(AlGaAs)包围而形成波导，对光信号具有很强的限制作用。在大多数光纤通信系统中都采用这种折射率导引的掩埋异质结(BH)半导体激光器。

2. DFB 和 DBR 激光器

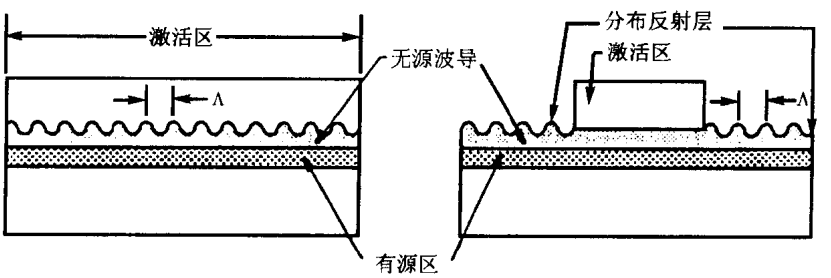
在前述的 F-P 谐振腔激光器中，两个端面是由解理或抛光制成的，以便获得产生激光所必须的光反馈。这种反馈也可以由激活区(光腔)内的布拉格光栅得到，相应的激光器称为分布反馈(DFB)激光器。如果布拉格光栅被做在激活区的两个端面之外，则构成分布布拉格反射(DBR)激光器。图 7.9 是 DFB 和 DBR 半导体激光器的示意图。人们用全息光刻或电子束光刻的技术，在激光器的有源层(或是相邻的波导层)中刻制一定周期、适宜深度的光栅，然后再继续外延生长第二种薄膜材料。由于两种材料的折射率不同，因而构成一系列周期性反射界面，组成了半导体介质光栅。无论是 DFB 或 DBR 激光器，激活区内电子和空穴复合产生的光子将受到每一条光栅的反射，从而形成光反馈。这十分类似晶体中布拉格衍射的情形。在晶体中，当光波以角度 θ 入射到一族晶面时，反射光发生相长干涉的条件(布拉格衍射条件)为

$$2d\sin\theta = m\lambda \qquad (m = 1,2,3,\cdots) \tag{7.81}$$

其中 d 为晶面间距，λ 为晶体中的光波波长。DFB 和 DBR 激光器的波纹光栅(布拉格光栅)类似于晶体的晶面族，光波传播方向与光栅垂直，即 $\theta=\pi/2$，故相长干涉条件为

$$\lambda_0 = \frac{2\bar{n}\Lambda}{m} \qquad (m = 1,2,3,\cdots) \tag{7.82}$$

Λ 为栅距；$\bar{n}$ 为介质的有效折射率；λ_0 为光波的真空波长，上式中用了 $\lambda=\lambda_0/\bar{n}$ 的关系；m 表示布拉格衍射的级数，$m=1$(一级布拉格衍射)时，两个相向传播的行波之间耦合最强。对于 $\lambda_0=1.55\,\mu$m 的激光器，取 $m=1$，$\bar{n}=3.3$，由式(7.82)得到 $\Lambda\simeq 0.235\,\mu$m。对于这种亚微米周期的光栅，可以采用全息技术或电子束技术得到。DFB 和 DBR 激光器产生激光振荡所必须的光反馈由布拉格光栅实现，而这种光栅可以通过平面工艺得到，所以它在集成光学中作为光源特别有用。集成光学使用在固体衬底上由平面工艺制造的小型光波导元件和电路。



(a) DFB　　(b) DBR

图 7.9　DFB 和 DBR 半导体激光器结构

7.4 量子阱(QW)激光器

QW 激光器的几何结构与通常的 DH 激光器类似，只是有源层很薄，其厚度约为 50～100Å，与受限电子的德布罗意波长(或电子自由程)相当，从而对载流子起着显著的量子限制作用。而在 DH 激光器中，有源层厚度约 1 000～3 000Å，虽然对载流子和光子起着限制作用，但它的电子学性质和光学性质仍然与体材料相同。由于量子限制作用，QW 激光器具有比较优越的性能。本节介绍这类激光器的一些基本知识。

7.4.1 量子阱中的载流子

图 7.10 表示 GaAs QW 激光器的分层结构和能带边，下面我们着重讨论量子阱导带内电子的运动。设垂直于阱壁的方向为 z 方向，则电子在平行于阱壁的平面(x-y 面)内是自由的，而在 z 方向的运动受到限制，故称为二维电子气。对于 $Al_xGa_{1-x}As/GaAs$ 异质结，导带边不连续量(势垒高度)$\Delta E_c \sim 1.25x(eV)$。为简单起见，我们把 ΔE_c 看作无限大(对于摩尔分数 $x \gtrsim 0.3$和阱宽 $L_z \gtrsim 100$Å 的情形，这是一个很好的近似)，即电子沿 z 方向是在一个无限深的势阱内运动，相应的电子波函数 $u(z)$满足薛定谔方程：

$$\frac{\hbar^2}{2m_n}u(z) + Eu(z) = 0 \tag{7.83}$$

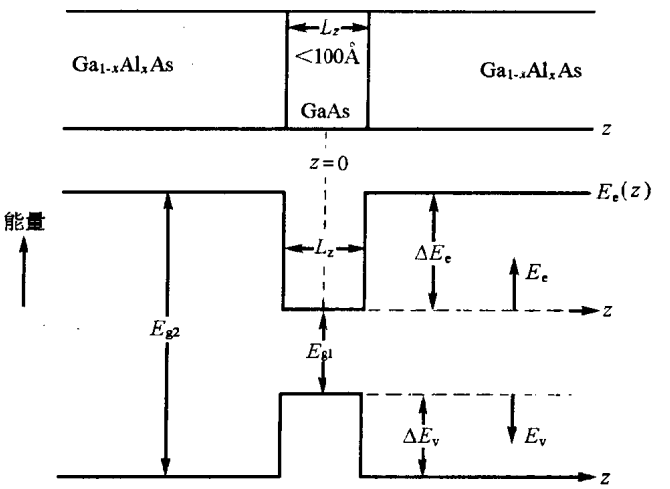


图 7.10 AlGaAs/GaAs/AlGaAs QW 激光器的分层结构和能带边

E 为电子沿 z 方向运动的能量，而限制电子运动的势能函数 $V(z)$可表成

$$V(z) = \begin{cases} 0 & (-L_z/2 < z < L_z/2) \\ \infty & (z < -L_z/2, z > L_z/2) \end{cases} \tag{7.84}$$

由于势垒无限高，电子不能逸出势阱，即在 $z = \pm L_z/2$ 处波函数 $u(z)$必须是零。利用这一条件可以得出，电子沿 z 方向运动的能量只能取下述的分立值：

$$E_{lc} = \frac{(\pi\hbar)^2}{2m_n}\left(\frac{l}{L_z}\right)^2 \quad l = 1,2,3,\cdots \tag{7.85}$$

相应的波函数为

$$u_l(z) = \begin{cases} \cos l\frac{\pi}{L_z}z, & l = 1,3,5,\cdots \\ \sin l\frac{\pi}{L_z}z, & l = 2,4,6,\cdots \end{cases} \tag{7.86}$$

另一方面，电子在 x-y 平面内自由运动，能量连续变化，

$$E_\perp = \frac{\hbar^2}{2m_n}(k_x^2 + k_y^2) = \frac{\hbar^2 k_\perp^2}{2m_n} \tag{7.87}$$

k_x 和 k_y 分别为 x 方向和 y 方向的电子波矢。所以，电子在势阱中运动的总能量为

$$E_c(k_\perp, l) = \frac{\hbar^2 k_\perp^2}{2m_n} + E_{lc} \tag{7.88}$$

导带底为能量零点。在上式中，l 取整数($l=1,2,3,\cdots$)，$k_\perp$ 在 0 至∞的范围内连续取值，电子的最低能量对应 $l=1$ 和 $k_\perp=0$ 的状态。l 一定而 $k_\perp$ 不同的所有能量状态构成一个子能带。

用空穴有效质量 m_p 代替电子有效质量 m_n，则上面的分析结果完全适用于量子阱中的空穴。这样，空穴在势阱中运动的总能量为

$$E_v = \frac{\hbar^2 k_\perp^2}{2m_p} + E_{lv}\ ,\ E_{lv} = \frac{(\pi\hbar^2)}{2m_p}\left(\frac{l}{L_z}\right)^2 \quad (l = 1,2,3,\cdots) \tag{7.89}$$

价带顶为能量零点，向下为空穴能量增加的方向，如图 7.10 所示。

能态密度

现在对于 $l=1$ 的子能带考虑电子的能态密度。为书写简单，令 $k = k_\perp$，但必须记住波矢 $\boldsymbol{k}$ 只有 x 方向和 y 方向两个分量，是二维情形。对于实际空间的单位面积，落入半径为 k、宽度 $\mathrm{d}k$ 的圆环内的状态数为

$$\rho(k)\mathrm{d}k = 2 \times \frac{2\pi k\,\mathrm{d}k}{(2\pi)^2} \tag{7.90}$$

因子 2 是因为每个 k 的本征值有两种自旋状态。$\rho(k)$为二维 $\boldsymbol{k}$ 空间的能态宽度，由式(7.90)有

$$\rho_{QW}(k) = k/\pi \tag{7.91}$$

在式(7.88)中令 $l=1$，即得相应情形下量子阱中电子总能量 E 和 k 的关系：

$$k = \sqrt{\frac{2m_n}{\hbar^2}(E - E_{1c})} \tag{7.92}$$

式中 $E_{1c} = \hbar^2\pi^2/2m_nL_Z^2$。由 $\rho(k)\mathrm{d}k = \rho(E)\mathrm{d}E$，得到量子阱中电子的能态密度

$$\rho_{QW}(E) = \frac{m_n}{\pi\hbar^2} \tag{7.93}$$

这是一个与能量无关的常数。

需要注意，上面的分析只考虑 $l=1$ 的子能带。当 $E > E_{2c}$时，电子既可以在 $l=1$ 的状态，也可以在 $l=2$ 的状态，故能态密度应为 $m_n/\pi\hbar^2$ 的 2 倍；在 E 超过 E_{3c}时应为 3 倍；等等。能级 E_{lc}每提高一级，总的能态密度增加 $m_n/\pi\hbar^2$，数学上表示为

$$\rho_{QW} = \sum_{l=1} \frac{m_n}{\pi\hbar^2}H(E - E_{lc})$$

其中函数 $H(x)$在 $x>0$ 时为 1，$x<0$ 时为 0。

图 7.11(a)表示了二维量子阱中电子和空穴的两个最低能级($l=1,2$)及其本征波函数；图 7.11(b)表示了量子阱中电子的能态密度分布(台阶状)，并画出了相应的体材料能态密度 ρ_{3D}(抛物线)作为比较。在抛物线能带结构中，由于靠近能带底部的态密度较小，处在这些能态上的载流子对受激发射不能产生有效的贡献，而在导带与价带的较高能态之间产生受激辐射复合之前又必须先填充这些能态，这必然会导致较高的阈值电流密度；换句话说，由于台阶状的能态密度分布，使量子阱激光器比起双异质结激光器来有较低的阈值电流密度。正如以前所指出的，由于这一原因以及量子阱层(有源层)很薄，使得量子阱激光器的阈值电流可比双异质结激光器低一个数量级以上，达到亚毫安水平。

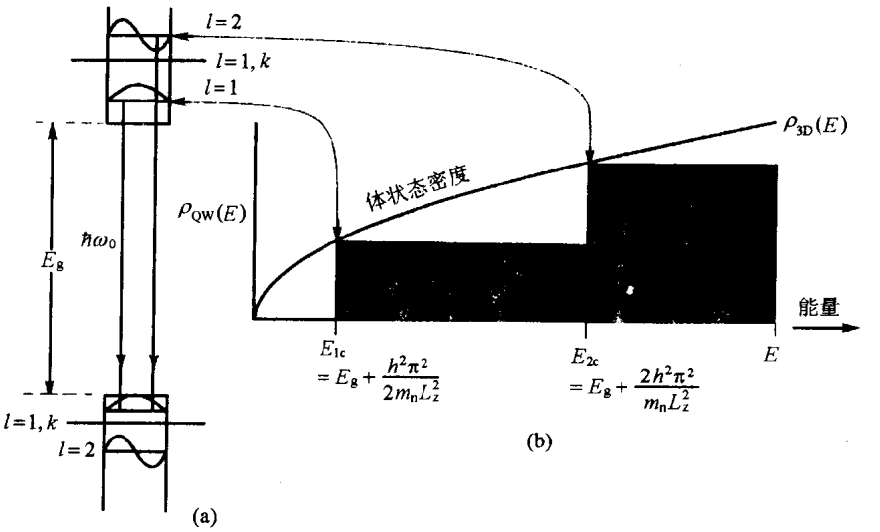


图 7.11 (a) 量子阱中电子和空穴的两个最低能态；(b) 能态密度分布

选择定则

由于能态被电子占据的几率(费米分布函数)从低能态到高能态减小，导带 $l=1$ 的能态容易被电子占据，价带 $l=1$ 的能态容易被空穴占据，所以量子阱的峰值光增益将来自 $l=1$ 的状态对之间的带间跃迁。我们就从这种跃迁着手，简短讨论量子阱中跃迁的选择定则。根据量子力学，跃迁速率(或光增益)正比于初、终态和极化量(x，y 或 z)三者相乘的积分的平方，而根据公式(7.86)，最低状态的电子和空穴的波函数正比于 $\cos(\pi z/L_z)$，由于

$$\int_{L_z/2}^{L_z/2} z\cos^2(\pi z/L_z)\mathrm{d}z = 0$$

所以光场 $\mathscr{E}$ 必须沿 x 或 y 方向偏振(即必须位于量子阱平面内)，沿 z 方向偏振的电场不能激发两个位置最低的能级之间的跃迁，从而不能引起光放大(或吸收)。这种光学跃迁是横电(TE)偏振。

7.4.2 量子阱中的光增益

这里，我们采用前面讨论体材料中光放大的方法，来得到量子阱的光增益系数。对于引起光放大的跃迁，导带电子态和价带空穴态(价带中没有被电子占据的能态)有相同的 l 值和 k

值，所以跃迁能量为

$$\hbar\omega = E_n - E_p = E_g + E_c(k,l) + E_v(k,l)$$
$$= E_g + \frac{\hbar^2}{2m_r}\left(k^2 + l^2\frac{\pi^2}{L_z^2}\right) \tag{7.94}$$

其中 m_r 是电子、空穴对的折合质量，

$$m_r = \frac{m_n m_p}{m_n + m_p} \tag{7.95}$$

先考虑 $l=1$ 这一对最低子能带之间跃迁引起的光增益。我们从表示体材料中光增益的公式(7.10)出发，只是将其中的 $g_v(\hbar\omega)$ 以 ρ_{QW}/L_z 代替(应当注意，推导增益系数时必须用反转载流子的体密度)。这样，有

$$G(\hbar\omega) = \frac{\hbar\lambda^2}{4\bar{n}^2\tau_r L_z}\rho_{QW}[f_n(E_n) + f_p(E_p) - 1] \tag{7.96}$$

将 $\rho_{QW} = m_r/\pi\hbar^2$ 代入上式，有

$$G(\hbar\omega) = \frac{m_r\lambda^2}{2h\bar{n}^2\tau_r Lz}[f_n(E_n) + f_p(E_p) - 1] \tag{7.97}$$

一般情形下，还应当考虑其他子能带($l=2,3,\cdots$)的贡献，这时只需做下述变换：

$$\frac{m_r}{\pi\hbar^2} \longrightarrow \frac{m_r}{\pi\hbar^2}\sum_{l=1} H(\hbar\omega - \hbar\omega_l)$$

其中 $\hbar\omega_l$ 是第 l 对子能带的带底之差。

现在讨论光增益 $G(\hbar\omega)$和占据几率 $f_n(E_n)$及 $f_p(E_p)$的关系。随着注入电流的增加，注入量子阱(有源区)的载流子浓度增加，准费米能级 E_{Fn}和 E_{Fp}分别向导带和价带深入，从而光频 ω 在较大范围内满足光放大的反转分布条件

$$f_n(E_n) + f_p(E_p) - 1 > 0$$

$f_n(E_n)$和 $f_p(E_p)$通过参与光跃迁的导带能态 E_n 及价带能态 E_p 与光频 ω 有关。对于能态密度分布为阶梯函数的理想情形，ω 的下限显然对应

$$\hbar\omega = E_g + \frac{\pi^2\hbar^2}{2mL_z} \tag{7.98}$$

即 $\hbar\omega$ 等于 $l=1$ 这一对子能带的带底($k=0$)之间的能量差。由于能态被占据的几率从低能态到高能态减小，这一频率的反转因子 $f_n(E_{1c}) + f_p(E_{1v}) - 1$ 为最大，即峰值增益 G_{max}由下式给出：

$$G_{max} = \frac{m_r\lambda^2}{2h\,\bar{n}^2\tau_r L_z}[f_n(E_{1c}) + f_p(E_{1v}) - 1] \tag{7.99}$$

E_{1c}和 E_{1v}分别为最低能量($l=1, k=0$)的导带子带边和价带子带边。所以，量子阱激光器的增益峰总是发生在最低能量子带边的跃迁，并不随注入电流改变(在有源区是体材料的双异质激光器中，增益峰的位置高于带边而强烈依赖于注入电流，或者说载流子浓度)；但随注入电流增加，它趋向于极限值 $m_r\lambda^2/(2h\,\bar{n}^2\tau_r L_z)$。

与体材料相比，量子阱的能态密度减少了一个维度的贡献，准费米能级 E_{Fn}和 E_{Fp}容易向各自的能带深入，亦即较低的注入电流密度(或载流子浓度)即可实现粒子数反转分布，以及载流子浓度较小变化将导致反转因子的较大变化。这样，与通常的双异质结激光器相比，量子阱

激光器由于透明载流子浓度 n_T 较小和有源区极薄而有较低的阈值电流密度，并且有较高的微分增益(dG_{max}/dn)。由前面关于激光器工作特性的讨论可知，微分增益高的激光器具有响应带宽(或响应速度)大和激光线宽小的优点。

进一步，我们着重考虑 $l=1$ 这一对能量最低的子能带，来讨论光增益和载流子浓度的关系。对于这种情形，不难证明(参见公式(4.76))，占据几率和载流子浓度 $n(=p)$ 的关系由下述公式表示：

$$f_n(E_{1c}) = 1 - e^{-n/N} \tag{7.100}$$

$$f_p(E_{1v}) = 1 - e^{-n/P} \tag{7.101}$$

其中，

$$N = \frac{4\pi m_n k_B T}{h^2 L_z} \quad , \quad P = \frac{4\pi m_p k_B T}{h^2 L_z} \tag{7.102}$$

结合方程(7.99)，(7.100)和(7.101)，得到

$$G_{max} = \frac{m_r \lambda^2}{2h\bar{n}^2 \tau_r L_z}[1 - e^{-n/N} - e^{-n/P}] \tag{7.103}$$

$$\frac{dG_{max}}{dn} = \frac{m_r \lambda^2}{2h\bar{n}^2 \tau_r L_z}\left[\frac{1}{N}e^{-n/N} + \frac{1}{P}e^{-n/P}\right] \tag{7.104}$$

显然，对于对称能带($m_n = m_p = m$)结构，正如双轴应变量子阱层的情形那样，$N=P$，方程(7.103)简化为

$$G_{max} = \frac{m_r \lambda^2}{2h\bar{n}^2 \tau_r L_z}[1 - 2e^{-n/N}] \tag{7.105}$$

透明载流子浓度定义为 $G_{max}=0$ 时所需的载流子浓度，这给出

$$n_T^S = N\ln 2 \tag{7.106}$$

对称能带结构量子阱透明载流子的面密度为

$$N_T^S = n_T^S L_z \tag{7.107}$$

结合方程(7.102)，(7.106)和(7.107)，得到

$$N_T^S = \frac{4\pi m k_B T}{h^2}\ln 2 \tag{7.108}$$

利用方程(7.104)，并假定透明时 $N=P$ 和 $2e^{-n/N}=1$，得到透明时的微分增益为

$$\frac{dG_{max}}{dN} = \frac{\lambda^2 \hbar}{(\ln 2)8\bar{n}^2 \tau_r k_B T} \tag{7.109}$$

7.4.3 应变层量子阱

构成量子阱的两种材料的晶格常数必须十分接近，以保证它们之间的晶格匹配，避免在其界面区产生大量失配位错而影响器件的性能。这就大大限制了量子阱结构材料选择的自由度。一个重大的进展是 1986 年提出了应变层量子阱的概念，它不仅开拓了量子阱材料选择的范围，而且展现出一系列优异特性。

应变层量子阱就是利用晶格不匹配的两种材料，用单原子层外延技术(MBE 或 MOCVD)生长在一起；由于生长层非常薄，从而使界面层保持在弹性限度之内，能够承受由于晶格失配

导致的内部应力，而保持一个经调整后的统一的晶格系统。在应变层量子阱结构中，阱材料将承受来自生长平面内的双轴应力，而发生应变。以 InP/$In_xGa_{1-x}As$/InP 量子阱结构为例，虽然在 $x=0.53$ 时 $In_xGa_{1-x}As$ 同 InP 晶格匹配，但双轴应变会来自 $x<0.53$ 时的张应变和 $x>0.53$时的压应变。当 $x<0.53$ 时，$In_xGa_{1-x}As$ 的晶格常数比 InP 的小，它的晶格在生长面(x,y)的方向受到拉伸，结果 z 方向(生长方向)的晶格常数减小；相反，$x>0.53$ 时，在生长平面内晶格受到压缩，z 方向的晶格常数会拉长。理论分析表明，双轴应变下的价带结构会发生变更，导带不连续性会增加。图 7.12 说明 $In_xGa_{1-x}As$ 在压应变和张应变下的能带结构。在无应变的情形下，重空穴(HH)带和轻空穴(LH)带在布里渊区中心简并，重空穴的有效质量比导带中电子的有效质量大得多，而轻空穴的有效质量和电子的相近。光学跃迁发生在导带和重空穴带之间。在应变下，由于双轴应力的作用，LH 带和 HH 带将发生分离，轻、重空穴的混合行为减弱。由图 7.12 可见，在压应变下，LH 带往下移动，光学跃迁将是横电(TE)偏振；而量子阱平面内 HH 带的形状变得陡峭，这意味着重空穴有效质量减小，对于典型的量子阱厚度，平面内有效质量估计是电子有效质量的 1.5 倍。这导致涉及光跃迁的导带和价带具有近似对称的能带结构。正是由于平面内有效质量的减小(z 方向仍保持原来的状态)，总的 HH 带态密度也将显著变小，从而使透明载流子浓度下降，使非辐射的俄歇复合和价带内的吸收下降。这些效应以及应变层内导带不连续的增加，使应变层量子阱激光器的阈值电流密度下降，量子效率提高。在张应变下，LH 带往上移动，光学跃迁是横磁(TM)偏振。对于较小的应变，能带混合效率使重空穴有效质量很大。当应变大于 1% 和阱宽也增大时，LH 带和 HH 带分开足够大，使混合效应降低，从而空穴有效质量下降，结果阈值电流密度显著下降。但对于更大的应变，阱宽度变得比维持弹性限度的临界厚度大，往往产生失配位错，使阈值电流密度增加。

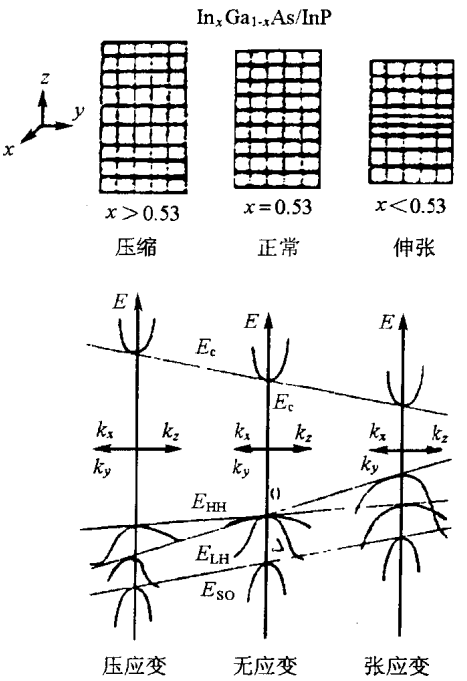


图 7.12　$In_xGa_{1-x}As$/InP 应变量子阱的能带(HH 为重空穴带，LH 为轻空穴带，SO 为自旋和轨道耦合分裂出的能带)

为了说明阈值电流密度下降，我们考虑导带和价带对称（$m_n = m_p = m$）的能带结构。对于对称结构的量子阱，透明载流子浓度（面密度）已由式(7.108)给出，为

$$N_T^S = \frac{4\pi m k_B T}{h^2}\ln 2 \tag{7.110}$$

而对于非对称能带结构（$m_p > m_n$）的量子阱，透明载流子面密度 N_T（$= n_T L_z$）可在方程(7.103)中令 $G_{max} = 0$ 得到，即 n_T 满足下述方程：

$$e^{-n_T/N} + e^{-n_T/P} = 1 \tag{7.111}$$

或

$$e^{-n_T/N}\left[1 + e^{n_T\left(\frac{1}{N}-\frac{1}{P}\right)}\right] = 1 \tag{7.112}$$

由于 $m_p > m_n$（即 $P > N$），方程(7.112)中的方括号是一个大于 2 的因子，故有

$$N_T > N_T^S = \frac{4\pi m k_B T}{h^2}\ln 2 \tag{7.113}$$

这说明比起对称能带结构来，非对称情形下必须有较大的 N_T 值，才能满足 $G_{max} = 0$ 的光透明条件。为了具体比较，对于非对称能带结构量子阱的透明载流子面密度，我们采用下述直观推理公式近似：

$$N_T = \left(\sqrt{\frac{m_p}{m_n}} + \frac{2m_p}{m_p + m_n}\right)\frac{N_T^S}{2} \tag{7.114}$$

设 $m_n = 0.041m_0$（m_0 为自由电子质量），$m_p = 0.70m_0$，则未应变量子阱激光器的透明载流子浓度比应变量子阱激光器的大 2 倍。

透明时的最大微分增益已由方程(7.109)给出。微分增益随电子和空穴的有效质量之间的非对称性下降而增加，在 $m_n = m_p$ 时最大。由于应变层量子阱激光器中微分增益增加，调制带宽增大，激光线宽减小。

顺便指出，Ge_xSi_{1-x}合金本征跃迁发光的波长范围是 1.3～1.55μm，这正是长距离光纤通信理想的波长窗口。但是，Ge_xSi_{1-x}合金是间接禁带半导体，不适合直接用作发光材料；而在 Ge_xSi_{1-x}/Si 的应变超晶格中，由于能带交叠，Ge_xSi_{1-x}变成直接禁带材料，提供了利用成熟的 Si 器件工艺制造半导体激光器的可能，也使大规模的光电子单片集成将能成为现实。

7.5 光电二极管

光电二极管是一类最重要的光探测器件，其最基本的结构是 p-n 结，工作时加反向偏压。受到光照时，光电二极管可以在很宽的范围内产生与入射光强度成正比的光生电流，所以能把光信号变成电信号，达到探测光信号的目的。在光纤通信系统中，实际使用的是 p-i-n 光电二极管和雪崩光电二极管，它们具有较高的量子效率和响应速度。

7.5.1 p-n 结的光电流

图 7.13 可以说明光电二极管的构造和基本原理。半导体吸收光之后，产生电子-空穴对。产生在耗尽层内或在耗尽层外载流子扩散长度内的电子-空穴对，最后将被电场（外加反向偏压引起的电场或自建电场）分开，它们漂移通过耗尽层，便在外电路产生电流。

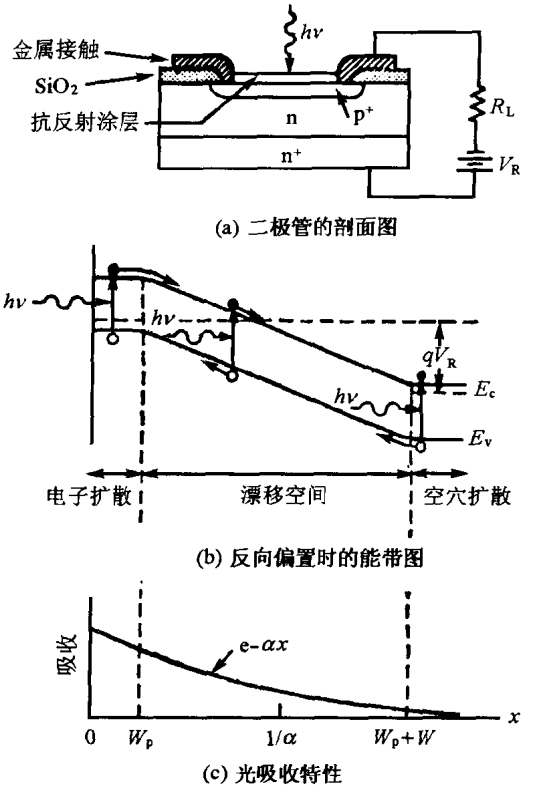


图 7.13　p-n 光电二极管的工作原理

假定热产生电流可以忽略，并且 p 型表面层很薄（$W_p \simeq 0$）以致其中的光吸收可以忽略。这样，电子-空穴对的产生率为

$$G(x) = \alpha \Phi_0 e^{-\alpha x} \tag{7.115}$$

Φ_0 是入射到受光面上的光子流密度，即 $x=0$ 处每秒每单位面积的光子数；α 为半导体的吸收系数。因此，耗尽层内产生的载流子所形成的漂移电流密度 J_{dr} 为

$$J_{dr} = q\int_0^W G(x)dx = -q\Phi_0(1-e^{-\alpha W}) \tag{7.116}$$

式中 W 为耗尽层宽度。

$x > W$ 时，半导体体内的少数载流子（空穴）密度由一维扩散方程决定：

$$D_p \frac{d^2 p_n}{dx^2} - \frac{p_n - p_{n0}}{\tau_p} + G(x) = 0 \tag{7.117}$$

式中，D_p 为空穴扩散系数，τ_p 为过剩空穴寿命，p_{n0} 为平衡空穴浓度。取 $x=\infty$ 时 $p_n = p_{n0}$ 和 $x=W$ 时 $p_n = 0$ 的边界条件后，方程(7.117)的解[①] 为

$$p_n(x) = p_{n0} - (p_{n0} + C_1 e^{-\alpha W})e^{(W-x)/L_p} + C_1 e^{-\alpha x} \tag{7.118}$$

① 方程(7.117)的标准形式为 $y'' - \frac{y'}{L_p^2} + \alpha \frac{\Phi_0}{D_p} e^{-\alpha x} = 0$，其解为 $y = Ae^{x/L_p} + Be^{x/L_p}$ + 特解，特解可令 $y = C_1 e^{-\alpha x}$ 代入微分方程，从而确定常数 C_1 以后得到。

$$C_1 = \left(\frac{\Phi_0}{D_p}\right)\frac{\alpha L_p^2}{1-\alpha^2 L_p^2} \tag{7.119}$$

式中，$L_p = \sqrt{D_p \tau_p}$。耗尽层以外的半导体(n层)内光生载流子扩散到耗尽层内形成的电流密度 J_{di}为

$$\begin{aligned} J_{di} &= qD_p\left(\frac{dp_n}{dx}\right)_{x=W} \\ &= -q\Phi_0\frac{\alpha L_p}{1+\alpha L_p}e^{-\alpha W} - qp_{n0}\frac{D_p}{L_p} \end{aligned} \tag{7.120}$$

由式(7.116)和(7.120)得到总的光电流密度为

$$J_L = -q\Phi_0\left(1-\frac{e^{-\alpha W}}{1+\alpha L_p}\right) - qp_{n0}\frac{D_p}{L_p} \tag{7.121}$$

在正常工作状态下，含 p_{n0}的项要小得多，可以忽略。同时，若受光面的反射率为 R_1(即实际上进入器件的光子部分为 $1-R_1$)，则光电流为

$$I_L = -qA\Phi_0(1-R_1)\left(1-\frac{e^{-\alpha W}}{1+\alpha L_p}\right) \tag{7.122}$$

前面的负号表示光电流方向由 n 侧流向 p 侧(反方向)。上式说明光电流与入射光强成正比。

表示光探测器效率的一个量是光电子流密度与入射光子流密度之比：

$$\eta_{探测} = \frac{I_L}{qA\Phi_0} = (1-R_1)\left(1-\frac{e^{-\alpha W}}{1+\alpha L_p}\right) \tag{7.123}$$

为了得到高的 $\eta_{探测}$，必须 R_1 小(在受光面上镀抗反射膜)和 W 大(采用 p-i-n 结构以增加耗尽层宽度)。

光照时 p-n 二极管的伏-安特性实际上为 p-n 结特性加上光电流。在忽略串联电阻和表面泄漏电流的理想情形下，其电流 I 为

$$I = I_s[e^{qV/mk_BT} - 1] - I_L \tag{7.124}$$

其中 I_s 是反向饱和电流；m 称为理想系数，是表示 p-n 结的特性参数，通常在 1～2 之间。作为光探测器，二极管外加反向偏压，将光信号转变为电信号；作为太阳电池，二极管不加偏压，将光能转换成电能。

1. 太阳电池

图 7.14 表示了一个单晶硅太阳电池的典型结构。它由表面上形成的浅 p-n 结、正面的指状条形欧姆接触、覆盖整个背面的欧姆接触以及正面的抗反射涂层构成。衬底用 p 型半导体，这是因为 p 型半导体的少数载流子(电子)的扩散长度比 n 型半导体的长的缘故。由于金属反射光，所以表面金属电极占据的面积越大，太阳电池的效率越低。但是，当电极面积越小时，因为电流流动时的电阻变大，效率也会降低。所以，必须将电极的宽度和电极的间隔设计成最佳值。

根据前述可知，p-n 结太阳电池的能带图基本上与反向偏压为零时的光电二极管的(参见图7.13)相同，其伏-安特性也由方程(7.124)描述。图7.15表示了光照时太阳电池的典型的伏-安特性。图 7.15 中，阴影面积是最大功率矩形($P_m = I_m V_m$)；I_{sc}是短路电流，$I_{sc} = I_L$；V_{oc}是开路($I=0$)电压，在式(7.124)中令 $I=0$，可得

$$V_{oc} = \frac{mk_BT}{q}\ln\left(1+\frac{I_L}{I_s}\right) \tag{7.125}$$

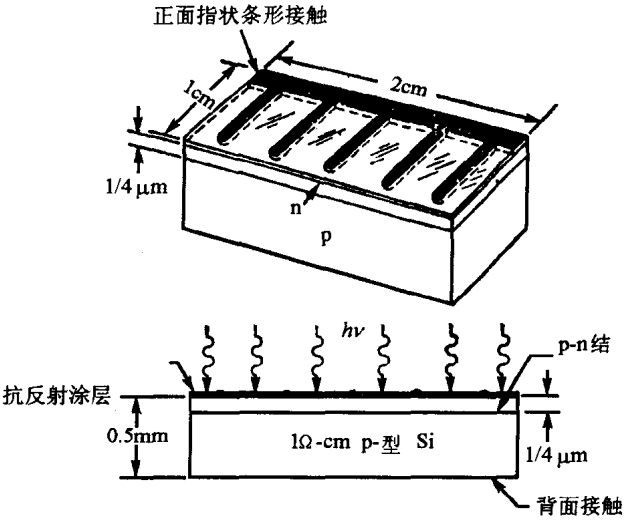


图 7.14　太阳电池的典型结构

上式表明,反向饱和电流越小和光电流越大,则开路电压越大。

太阳电池的输出功率为

$$P = VI = V[I_L - I_s(e^{qV/mk_BT} - 1)] \tag{7.126}$$

当 $dP/dV=0$ 时,得到最大功率条件:

$$V_m = \frac{1}{\beta}\ln\left[\frac{1+(I_L/I_s)}{1+\beta V_m}\right]$$

$$I_m = \frac{\beta V_m(I_L+I_s)}{1+\beta V_m}$$

式中 $\beta = q/mk_BT$。

太阳电池接上负载可得到功率输出,也就是说太阳能发电。以 R 表示负载电阻,则图7.15所示的光照时的伏-安特性与负载线的交点为工作点,负载消耗的功率随负载电阻值而变。当连接的负载使功率输出为最大值 $P_m(=V_mI_m)$时,P_m 与入射太阳光的功率 P_{in} 之比称为转换效率

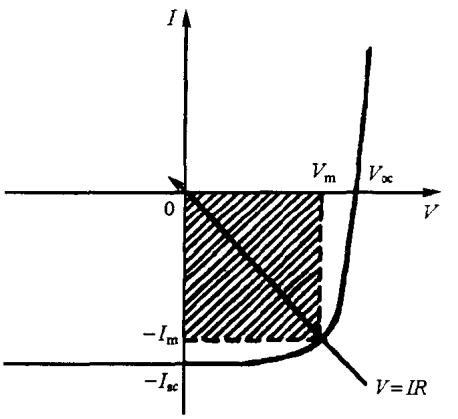


图 7.15　光照下太阳电池的伏-安特性

$$\eta = \frac{V_mI_m}{P_{in}} = \frac{FFI_LV_{oc}}{P_{in}} \tag{7.127}$$

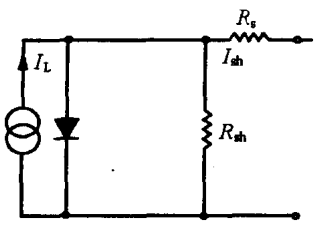


图 7.16　太阳电池的等效电路

式中 FF 是填充因子,定义为

$$FF = \frac{V_mI_m}{V_{oc}I_L} \tag{7.128}$$

对于给定的大气质量(AM)状态,短路电流 I_L 是电子电荷 q 和太阳能光谱内 $h\nu \geqslant E_g$(E_g 是半导体的禁带宽度)的有用光子数的乘积,输入光功率 P_{in}是太阳能光谱中所有光子的积分。

有许多因素造成太阳光中的能量损失,影响太阳电池的转

换效率。首先是正面欧姆串联电阻和 p-n 结复合电流(泄漏电流)引起的损耗,等效电路如图 7.16 所示。图中,串联电阻 R_s 由电极等引起,并联电阻 R_{sh} 相当于阻止 p-n 结产生泄漏电流的电阻,恒流源 I_L 相当于与光强度成比例的二极管反向电流。由等效电路图可以得出太阳电池两端的电流和电压的关系为

$$I = I_L - I_s\left[\exp\left\{\frac{q(V + R_sI)}{mk_BT}\right\} - 1\right] - \frac{V + R_sI}{R_{sh}} \tag{7.129}$$

为了使太阳电池输出更大的功率,必须尽量减小串联电阻,增大并联电阻;也就是说,必须尽量降低表面的反射,减小由于表面以及缺陷引起的复合,优化电极结构。

在太阳光的能量损失中,还有一部分是不可避免的,这就是:$h\nu < E_g$ 的光不能被半导体吸收而透射出去所产生的损失;$h\nu > E_g$ 的光只能取出能量与禁带宽度 E_g 相当的部分,而有 $h\nu - E_g$ 的能量通过放出声子而转换成热能。由于太阳电池的材料决定了有不可避免的损失,所以为获得高效率的太阳电池,必须使太阳电池的光谱响应与太阳光谱尽量一致。制造太阳电池的材料有单晶硅、非晶硅、化合物半导体等,在 AM1.5 的条件下,单晶硅太阳电池的效率达到 24.0%,非晶硅太阳电池为 13.2%,InGaP/GaAs 叠层太阳电池达到 30.28%。

7.5.2 p-i-n 光电二极管和雪崩光电二极管

1. p-i-n 光电二极管

在 p 层和 n 层中间夹入一层本征(或低掺杂)的 i 层,以增加耗尽层的宽度,这种结构的光电二极管称为 *p-i-n* 光电二极管。*p-i-n* 光电二极管是常用的光探测器之一。

由于器件反向偏置,暗态电流为二极管反向饱和电流 I_s,与偏压无关;光电流 I_L 主要由耗尽层(i 层)产生的载流子引起。假定 i 层比电子或空穴的扩散长度厚得多,则根据式(7.122)可得

$$I_L = qA\Phi_0(1 - e^{-\alpha W})(1 - R_1) \tag{7.130}$$

Φ_0 是入射的光子流密度,W 是耗尽层宽度。因为 W 越大,i 层越能吸收更多的光能,所以能够得到更大的光电流,即更高的响应度。但是,如果 W 太大,载流子的渡越时间就会很长,从而降低器件的响应速度。对于高速器件,W 为微米量级甚至更小。

在长距离光纤通信系统中,p-i-n 光电二极管常采用 p-InP/i-InGaAs/n-InP 的双异质结结构。InP 的禁带宽度为1.35eV,对波长大于 0.92μm 的光不吸收;而 InGaAs 的禁带宽度为 0.75eV(对应截止波长 1.65μm),在 1.3～1.6μm 波段上表现出较强的吸收,几微米厚的 i 层就可以获得很高的响应度。这样,对于光纤通信的低损耗波段,光吸收只发生在 i 层,完全消除了扩散电流的影响。

2. 雪崩光电二极管(APD)

APD 也是最常用的一类光探测器,它利用器件内部的雪崩倍增过程使光吸收产生的电流得到放大。通常的 p-i-n 光电二极管没有这种放大作用。

图 7.17 表示 APD 的典型结构和反向偏压 V_R 下的电场分布。V_R 足够大时,n^+-p 结的强电场将引起雪崩倍增,在这里形成"倍增区",使光生电流放大;i 层的厚度比较大(直接禁带材料为微米量级,间接禁带材料为几十微米),是光吸收的区域(吸收区),其间的电场不足以发生碰撞电离,但通常能使载流子以饱和漂移速度 v_s 运动。n 型保护环的作用是防止 n^+-p 结的边缘击穿,注意到中间 n^+-p 结将先于 n-p 结保护环击穿就不难理解这一点。

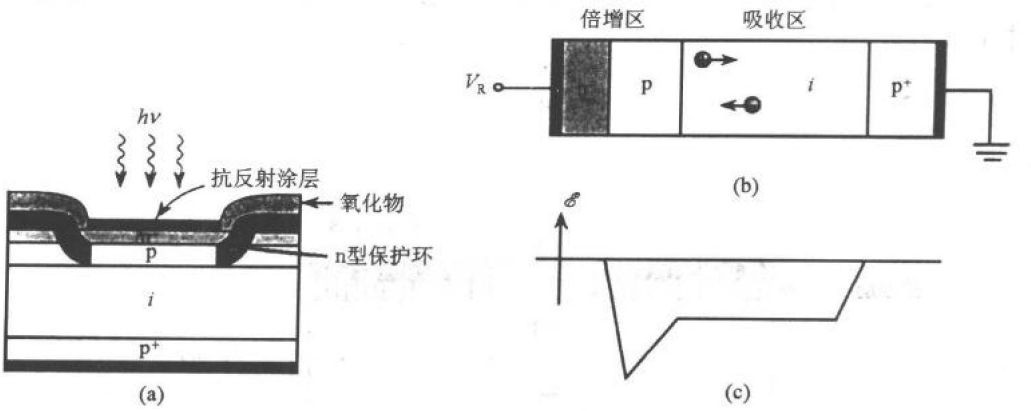


图 7.17　APD 的典型结构

假设雪崩过程在厚度为 d 的倍增区上 $x=0$ 处开始并且 $x=d$ 处只有电子注入，即 $I_n(d)=I$，I 是通过器件的总电流。这样，根据在 2.6 节中的讨论，器件的直流倍增因子 $M_0=I_n(d)/I_n(0)$可表示为

$$M_0=\frac{1}{1-\int_0^d \alpha_n\left[e^{-\int_0^x(\alpha_n-\alpha_p)dx'}\right]dx}$$

$$=\frac{1-\alpha_p/\alpha_n}{e^{-(\alpha_n-\alpha_p)d}-\alpha_p/\alpha_n} \tag{7.131}$$

上式的后一步假设了 α_n，α_p 与位置 x 无关。由上式可知，倍增因子与碰撞电离率的比值 α_p/α_n有很大关系，当 $\alpha_p=0$ 时(只有电子参加雪崩倍增过程)，有

$$M_0=e^{\alpha_n d}$$

倍增因子随倍增区的厚度 d 指数增加；如果 α_n 和 α_p 相同，则有

$$M_0=\frac{1}{1-\alpha_n d}$$

击穿电压相当于 $\alpha_n d=1$ 的情形。尽管在 α_n 和 α_p 相近的情形下可以从很窄的倍增区获得较高的增益，但这时 APD 的噪声很大，且响应速度(带宽)大大减小。所以 APD 的倍增区常采用 α_n 和 α_p 差别很大的材料。

本章的叙述一再说明Ⅲ-Ⅴ族化合物半导体是制造光电器件的重要材料。原则上，在直接禁带半导体的情形下，光探测器的吸收区不必很长，APD 可以由很薄的吸收区和倍增区构成。但是，对于窄禁带材料(例如 InGaAs 的禁带宽度只有 0.75eV)，碰撞电离所需的强电场会引起很大的来自带间隧道效应的漏电流。为了避免这种现象，通常采用分别吸收和倍增(简写为 SAM)的 APD 结构，光吸收发生在窄带材料内，雪崩倍增过程发生在宽禁带材料内。图 7.18 表示 1.5～1.65μm 波长的光通信用 InGaAs APD 的典型结构，它是在 n^+-InP 衬底上外延生长 n^--InGaAs 光吸收层和 n^--InP 倍增层并在这两层中间夹入 n^--InGaAsP 层制成的。夹入 InGaAsP 层是用来缓和 InGaAs 层和 InP 层间的禁带宽度 E_g(分别为 0.75eV 和 1.35eV)的差异所产生的价带不连续，使光吸收层产生的空穴迅速流入倍增层。作为倍增层的 InP 材料的

空穴电离率大于电子电离率($\alpha_p > \alpha_n$),所以这种 APD 的设计是倍增过程由空穴碰撞而在 n 型 InP 中形成。

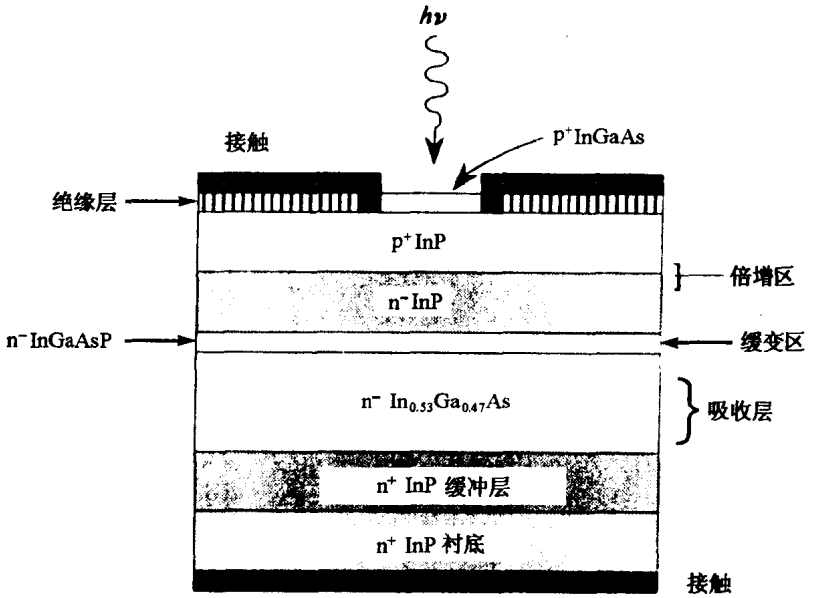


图 7.18

7.5.3 光电二极管的特性参数

1. 量子效率和响应度

量子效率是每个入射光子产生的电子-空穴对数目:

$$\eta = \frac{I_L/q}{P_0/h\nu} \tag{7.132}$$

I_L 是吸收波长为 λ(相应的光子能量为 $h\nu$)、功率为 P_0 的入射光产生的光电流,P_0 和入射光子流密度 Φ_0 的关系为 $P_0 = A(1-R_1)\Phi_0 h\nu$。于是,利用式(7.130)可得

$$\eta = 1 - e^{-\alpha W} \tag{7.133}$$

η 和吸收系数 α 的关系十分密切。图 7.19 表示了一些半导体的光吸收系数随波长的变化情况。$\alpha = 0$ 时的波长 λ_c 称为截止波长,这对应光子能量 $h\nu = E_g$,

$$\lambda_c = \frac{1.24}{E_g}[\mu m] \tag{7.134}$$

光电二极管只能探测 $\lambda < \lambda_c$(或 $h\nu > E_g$)的光辐射。光响应也有短波限,这是因为波长很短时,α 值很大,大部分辐射在表面附近被吸收,而表面的复合时间又很短,因此光生载流子在被 p-n 结收集之前就已经被复合掉了。

响应度表征光电二极管的转换效率,定义为光电流与光功率之比:

$$R = \frac{I_L}{P_0} \tag{7.135}$$

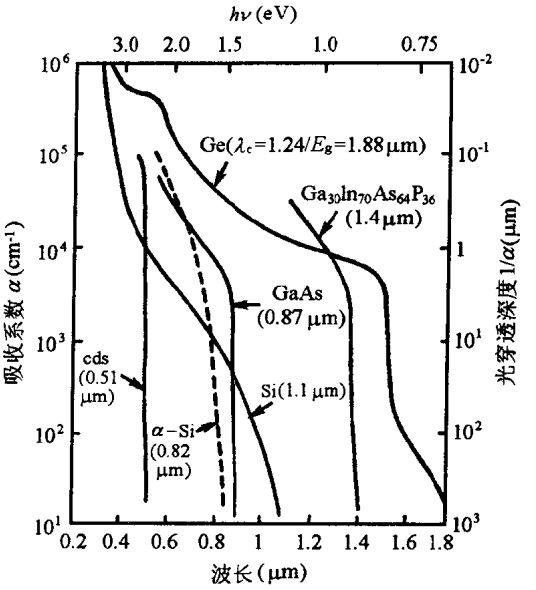


图 7.19

用量子效率 η，有

$$R = \eta \frac{q}{h\nu} = \frac{\eta\lambda(\mu m)}{1.24} \quad [A/W] \tag{7.136}$$

2. 响应速度(带宽)

在光纤通信系统中，要求光电二极管对入射光强度受到高频调制引起的变化快速响应。载流子渡越耗尽层的时间 t_r 将限制响应频率(即响应带宽)。分析表明，当调制频率 ω 满足下述关系时，交流光电流的大小下降到低频时的 $1/\sqrt{2}$ 倍：

$$\omega t_r = 2.4 \tag{7.137}$$

这一频率称为 3dB 频率，或 3dB 带宽，我们以 f_{3dB} 表示，它与耗尽层宽度 W 和饱和漂移速率 v_s 的关系为

$$f_{3dB} = \frac{2.4}{2\pi t_r} = 0.44 \frac{v_s}{W} \tag{7.138}$$

显然，减小耗尽层宽度，器件的响应带宽将增加；但另一方面，器件的响应度因吸收光子较少而下降。

对于 APD，设雪崩过程由电子注入倍增区引起，则产生的电子与原始电子一道通过倍增区，而产生的空穴朝相反的方向漂移。如果这些空穴不太可能产生新的电子(即 $\alpha_p \to 0$)，则所有电荷移出耗尽层的时间等于电子和空穴的渡越时间之和，带宽和倍增因子无关。但是，如果初始雪崩产生的空穴在倍增区内产生电子-空穴对的几率显著，则或多或少有新的电子产生。这一过程是再生雪崩过程，它导致原始电子在扫出倍增区以后在该区长时间内存在电子，从而引起附加的渡越时间延迟。倍增因子越大，附加的时间越长。由于这个时间的存在，在频率 ω 的高频调制光入射下，APD 的增益将会下降，从而形成对器件响应带宽的限制。分析表明，器件的高频增益可以写成

$$M(\omega) = \frac{M_0}{[1 + (\omega t_1 M_0)^2]^{1/2}} \tag{7.139}$$

式中，M_0 为 APD 的直流倍增因子；t_1 为等效渡越时间，在 $\alpha_n > \alpha_p$ 的情形下，$t_1 \simeq (\alpha_p/\alpha_n)t_r$。由式(7.139)可得 APD 的 3dB 带宽为

$$f_{3dB} = (2\pi t_1 M_0)^{-1} \tag{7.140}$$

上式表明了器件的增益-带宽乘积为常数，如果想提高探测微弱光信号的灵敏度，就必然降低器件的响应带宽；也表明了采用 $\alpha_p \ll \alpha_n$(或 $\alpha_n \ll \alpha_p$)的材料制作 APD，可望获得较高的响应带宽。

实际上，光电二极管的响应速度往往受限于结电容 C 和负载电阻 R 所决定的时间常数 RC。在这种情形下，最高响应频率(带宽)可表示为

$$f = \frac{1}{2\pi RC} \tag{7.141}$$

所以，为了提高响应带宽，应尽量减小结电容，尤其是尽量减小结面积。

3. 噪声性能

噪声是信号上附加的无规则起伏，从而信号变得模糊甚至被淹没。所以，噪声是通信中的一种基本限制。在光电二极管中，主要有散粒噪声和热噪声，下面讨论这两类噪声。

散粒噪声是由一个个光子产生的不均匀的或杂乱的电子-空穴对引起的，也就是说由通过器件的粒子(电子或空穴)数无规则起伏引起的。假定粒子数偏差服从高斯分布，则当时间间隔 Δt 内通过的平均粒子数为 $\overline{N}(=a\Delta t$，a 为粒子流强度)时，在 Δt 内得到 N 个粒子($N \gg 1$)的几率为

$$\begin{aligned} P(N, \Delta t) &= \frac{1}{\sqrt{2\pi(a\Delta t)}} e^{-\frac{(N-a\Delta t)^2}{2a\Delta t}} \\ &= \frac{1}{\sqrt{2\pi\overline{N}}} e^{-\frac{\Delta N^2}{2\overline{N}}} \end{aligned} \tag{7.142}$$

其中 $\Delta N = (N - \overline{N})$是偏离平均值的起伏。当 $N = \overline{N}$ 时，该几率分布函数有最大值。根据统计学原理和利用式(7.142)，不难得到粒子数的均方根偏差(噪声)为

$$\sqrt{\overline{(\Delta N)^2}} = \sqrt{\overline{(N - \overline{N})^2}} = \sqrt{\overline{N}} \tag{7.143}$$

即

$$\sqrt{\overline{(\Delta N)^2}} = \sqrt{a\Delta t} \tag{7.144}$$

在式(7.144)中，a 为粒子流强度，它与电流强度 I 的关系为

$$a = \frac{I}{q} \tag{7.145}$$

Δt 是观测时间，以 Δf 表示测量的频率范围(带宽)，则有

$$\Delta f \simeq \frac{1}{2\Delta t} \tag{7.146}$$

从式(7.144)，(7.145)和(7.146)可以得到探测器电流的均方根噪声即散粒噪声电流 i_{ns} 为：

$$i_{ns} = q\frac{\sqrt{\overline{(\Delta N)^2}}}{\Delta t} = \sqrt{2qI\Delta f} \tag{7.147}$$

通常写成

$$i_{ns}^2 = 2qI\Delta f \tag{7.148}$$

热噪声来自阻值为 R 的电阻体发出的电磁辐射的低频部分，由载流子无规则散射引起。实际上，任何一个温度高于绝对零度的物体都要发出这种噪声。根据普朗克的黑体辐射理论，热噪声的功率是

$$P = \left(\frac{h\nu}{e^{h\nu/k_B T} - 1}\right)\Delta f \tag{7.149}$$

式中，括号内的项是频率为 ν 的电磁振动方式的平均能量，Δf 是探测用的频带宽度。在电磁辐射的高频部分，$k_B T$ 和 $h\nu$ 比起来很小，热噪声的功率很小。所以，在加有偏压的光电二极管中，主要是因入射光无规则地产生载流子而引起的散粒噪声，热噪声并不重要。但是，在探测装置中设置有放大器的场合下，热噪声是放大器的主要噪声，它将限制光探测器可探测到的最小光信号功率(即探测器的灵敏度)。

在电磁辐射的低频部分，即对微波高频或任何更低的频率，除人工产生的极低温度外，式(7.149)可近似为

$$P = k_B T\Delta f \tag{7.150}$$

相应的热噪声电流(均方值)为

$$i_{nt}^2 = \frac{4k_B T\Delta f}{R} \tag{7.151}$$

将式(7.148)和式(7.151)相加，就得到光电二极管在接有输入电阻为 R 的放大器时的噪声电流 i_n^2(均方值)为：

$$i_n^2 = 2q(I_L + I_D)\Delta f + \frac{4k_B T\Delta f}{R} \tag{7.152}$$

其中，I_L 是入射光在光吸收层中产生的电流，即信号电流；I_D 是在光吸收层中与入射光无关的、由热激发的载流子引起的电流，即暗电流。一般的光电二极管的串联电阻成分 R_s 可忽略不计，并联电阻成分 R_{sh} 很大。R 等于 R_{sh}、负载电阻 R_L 以及接在负载电阻上的放大器的有效输入电阻 R_A 的并联电阻值。另外，T 一般表示放大器的噪声系数和绝对温度的乘积，即有效温度。

利用式(7.132)，可将光电二极管的信号电流 i_s 简单地表示为(令 $i_s = I_L$)

$$i_s = \frac{q\eta P_0}{h\nu} \tag{7.153}$$

η 为量子效率，$h\nu$ 为入射光的光子能量，P_0 为入射光的光功率。在负载电阻 R 两端产生的信号功率 P_s 为：

$$P_s = i_s^2 R \tag{7.154}$$

信噪比(S/N)是信号功率 P_s 与噪声功率 $i_n^2 R$ 之比，即

$$\frac{S}{N} = \frac{(q\eta P_0/h\nu)}{2q(q\eta P_0/h\nu + I_D)\Delta f + 4k_B T\Delta f/R} \tag{7.155}$$

在忽略暗电流和热噪声的条件下，光电二极管的信噪比可通过式(7.155)得到，为

$$\frac{S}{N} = \frac{\eta P_0}{2h\nu\Delta f} \tag{7.156}$$

噪声等效功率(NEP)定义为产生与探测器噪声输出大小相等的信号所需要的入射光功率。NEP 标志探测器可探测的最小光功率。在式(7.156)中令 $S/N=1$,就得到光电二极管的 NEP 为:

$$\mathrm{NEP}=\frac{2h\nu\Delta f}{\eta} \tag{7.157}$$

探测率 D 定义为:

$$D=\frac{1}{\mathrm{NEP}} \tag{7.158}$$

D 依赖于探测器的面积 A 和带宽 Δf。为了排除这些变量的影响,引入比探测率 D^*:

$$D^*=\frac{(A\Delta f)^{1/2}}{\mathrm{NEP}} \tag{7.159}$$

D^* 是探测器的常用优值。选用探测器时,一旦带宽条件满足,就应当选用 D^* 值高的器件。

对于雪崩器件,还应考虑雪崩噪声,这种噪声是由雪崩倍增过程的随机性产生的。在 APD 中,光生载流子注入倍增区,不仅注入的光电流被倍增,而且噪声也被放大。这样,表示器件散粒噪声的公式(7.148)应被下式取代:

$$i_{\mathrm{ns}}^2=2qIM^2F\Delta f \tag{7.160}$$

式中,M 为倍增因子,F 称为过剩噪声因子。在电场均匀的情形下,F 可写作:

$$F=M\left[1-(1-k)\left(\frac{M-1}{M}\right)^2\right] \tag{7.161}$$

当 M 很大时,上式近似为

$$F=2(1-k)+kM \tag{7.162}$$

其中,$k=\alpha_{\mathrm{p}}/\alpha_{\mathrm{n}}$ 或 $\alpha_{\mathrm{n}}/\alpha_{\mathrm{p}}$,前者对应雪崩过程由电子注入倍增区引起,后者对应由空穴引起。式(7.162)表明,为了减小雪崩噪声,应当选用 α_{n} 和 α_{p} 相差很大的半导体,这和前面谈到的提高器件的响应速度时对材料的要求是一致的。硅的 α_{n} 比 α_{p} 大得多,因此硅 APD 具有良好的性能,在 0.85μm 光谱区的接收装置中多数使用这种器件。此外,由于散粒噪声随 M 迅速增加,APD 的内部增益不能过大。

主要参考文献

[1] 半导体物理学(上、下册).叶良修编著.北京：高等教育出版社,1983

[2] 半导体物理基础.黄昆、韩汝琦著.北京：科学出版社,1979 年

[3] 固体物理学.黄昆原著、韩汝琦改编.北京：高等教育出版社,1988 年

[4] 半导体器件:物理和工艺(*Semiconductor Devices: Physics and Technology*)[美]S. M. Sze 著.王阳元、嵇光大、卢文豪译.北京：科学出版社,1992 年

[5] 半导体器件物理(*Physics of Semiconductor Devices, 2nd Ed.*)[美]S. M. Sze 著.黄振岗译.北京：电子工业出版社,1989 年

[6] 王阳元文集. 王阳元著.北京：北京大学出版社,1998 年

[7] *Physics of Semiconductor Devices*. M. Shur. Prentice-Hall, Englewood Cliffs. 1990

[8] *Fundamentals of Semiconductor Theory and Device Physics*. Prentice-Hall, Englewood Cliffs. 1989

[9] *High-Speed Semiconductor Devices*. Edited by S. M. Sze Wiley, 1990

[10] Modern Semiconductor Devices Physics. [美] *S. M. Sze* 主编. 刘晓彦、贾霖、康晋锋译,北京：科学出版社,2001

[11] 多晶硅发射极晶体管和集成电路.王阳元、张利春、赵宝瑛编译.北京：科学出版社,1993 年

[12] 用于 VLSI 模拟的小尺寸 MOS 器件模型:理论与实践(*MOSFET Models for VLSI Circuit Simulation Theory and Practice*).[奥]N. Arora 著.张兴、李映雪等译,原著.北京:科学出版社,1999 年

[13] MOS 晶体管的工作原理及建模(*Operation and Modling of The MOS Transistor*).[美]Y. P. Tsividis 著.叶金官等译.西安：西安交通大学出版社,1989 年

[14] SOI 技术(*Silicon-On-Insulatior Technology*).[美]J. P. Colinge 著.武国英等译.北京：科学出版社,1993 年

[15] 热电子微波发生器(*Hot Electron Microwave Generation*).[美]J. E. Carroll 著.赵学恕、汪兆平、武振民、韩忠惠译.北京：科学出版社,1975 年

[16] 介观物理.阎守胜、甘子钊主编.北京大学出版社,1995 年

[17] *Optoelectronics: An Introduction to Materials and Devices*. J. Singh. McGraw-Hill, 1996

[18] *Optical Electronics*, 4th ed A. Yariv Holt, Rinehard and Winston. 1991

[19] K. Suzuki, *IEEE Trans. Electron Devices*, 1991, ED-38, 1868

[20] K. Lee, M. Shur, T. Drummond and H. Mor koc, *IEEE Trans. Elec. tron. Devices*, 1983, ED-30, 207

[21] A. Ejeldly and M. Shur, *IEEE Trans. Electron Devices*, 1993, ED-40, 137